

# Integrated Image Recommendation Based on Category and Color Utilizing CNN and VGG-16



R. Bhuvanya and M. Kavitha

**Abstract** In earlier days, image retrieval is based on text whereas due to the improvement in technology, and this paper aims at deep learning (DL) techniques to produce integrated image recommendations. As the recommendation system has become inevitable in e-commerce sites, the proposed work is based on recommending the matching color of the same category product. Initially, the preprocessing of the dataset is carried out by applying lemmatization and stop words removal then by applying term frequency-inverse document frequency (TF-IDF), the top 15 frequent categories of the product are identified which is considered for further processing. Then, the Web scraping technique is applied to retrieve the images from the product URL. To identify the color of the product, the color histogram approach is incorporated. The fetched images are fed into the DL techniques of convolutional neural network (CNN) and VGG-16 to recommend the matching color of the same category product. The evaluation metrics of precision, recall, F1 score, and accuracy are applied to evaluate the implemented model. From the experimented results, it is justified that the VGG-16 technique scores better than CNN.

**Keywords** Color histogram · Convolutional neural network · Image retrieval · TF-IDF · VGG-16

## 1 Introduction

As the volume of data grows rapidly, it is essential to improve the quality of recommendations. Hence, the data scientists have replaced the traditional machine learning (ML) algorithms with deep learning (DL) models. In the past few years, deep learning techniques have perceived great success in computer vision applications. Recently,

---

R. Bhuvanya (✉) · M. Kavitha  
Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, Chennai, India  
e-mail: [bhuvanyaraghunathan@gmail.com](mailto:bhuvanyaraghunathan@gmail.com)

M. Kavitha  
e-mail: [kavitha@veltech.edu.in](mailto:kavitha@veltech.edu.in)

the DL models have started their influence in the field of recommender systems (RS) as it achieves high-quality recommendations. Before diving into the techniques of DL, this introductory session explains the basic concept of RS. Recommender system analyzes the user and item profile [1] and helps to discover new items by reducing the searching time of the user. It is usually classified into collaborative, content-based, and hybrid filtering techniques [2]. Collaborative filtering generates the recommendation list by considering the user-item historical interactions, browser history, and user ratings whereas content-based filtering [3] generates the recommendation list by considering the features of items such as color, shape, text, etc. The hybrid filtering produces the recommendation by integrating one or more filtering techniques [4]. The category of recommendation can be non-personalized, semi-personalized, and personalized [5]. If the recommendation list is generated based on the relevance of the most popular category, it is considered a non-personalized system. If the recommendation is generated based on a group, for instance, students, professionals, old aged people, etc., then it is a semi-personalized recommendation. At this level, the same category of people will get the same type of recommendations. The last level of personalized recommendation provides the items by analyzing the current user of the system. The generated list will be made only to the specific user. This recommendation can be applied to various application domains such as e-commerce, healthcare, tourism, and education. The famous e-commerce site Amazon.com has launched its recommendation engine two decades ago and millions of users have benefitted from the RS, by discovering unknown items [6, 7].

Deep learning is a subset of ML which is a subset of artificial intelligence (AI) [8]. With the ability to work with a large set of data along with good computational power, DL algorithms can able to self-learn the hidden patterns to make predictions and recommendations. The difference between the prediction and recommendation is the predictions help to quantify the items whereas the recommendations help the users to discover the unknown items. With the increase of research publications based on deep learning-based recommendations, it is evident that the DL models have proved their efficiency. All the social media sites and e-commerce sites have implemented the DL models for recommending items. For instance, YouTube, e-Bay uses deep neural network (DNN) models, and Spotify, Netflix uses convolutional neural network. When the neural network is applied, it can process complex user-item interaction patterns. Also, due to the advancement of deep learning models, it is possible to build hybrid models for recommending items. The deep learning models proved their efficiency in the tasks of both supervised and unsupervised ML. As the field is teeming with new ideas, DL models have become more successful in image processing and speech recognition techniques, e-commerce, healthcare, advertising, entertainment, etc. The major reasons for implementing DL models in RS are non-linear transformation, representation learning, sequence modeling, and flexibility [9].

**Non-linear Transformation:** The traditional matrix factorization technique combines the user and item latent factors linearly. But the neural networks (NN) are capable of approximate the function by applying various activation functions.

**Representation Learning:** The real-world applications usually contain a large amount of descriptive information about user-item interactions. To get a better quality recommendation, it is essential to make use of this information effectively. When the deep learning model is incorporated, it learns the features from the raw data. Also, it can make use of text, image, audio, and video to build a recommendation model.

**Sequence Modeling:** In the past decade, DNN has exposed its efficient results in natural language processing and speech recognition techniques. The deep learning techniques of Recurrent Neural Network (RNN) and Convolutional Neural Network (CNN) work effectively in the above-mentioned tasks. RNN is a feedforward network, in which the information moves in the forward direction. It consists of input, hidden, and output layers. RNN achieves this task by the presence of internal memory. Because of the internal memory, the RNN can make the prediction easier, hence it is useful for the time series data, sentiment analysis. CNN works on the image by taking a segment also known as filters. Sequence modeling has become important in session-based recommendations. Hence, it can be concluded that the deep learning models are extremely useful for the pattern mining task.

**Flexibility:** Deep learning has the framework of Tensorflow, PyTorch, DeepLearning4j, Keras, ONNX, Caffe, etc. Hence, it provides an easy way to combine various neural network models which makes the task of prediction and recommendation efficient.

This paper focuses on color and category-based image retrieval [10, 11] by implementing convolutional neural network (CNN) and VGG-16. Initially, the category of the product needs to be identified, which is done by applying natural language processing techniques. Then, the color of the user interested product is identified through color histogram which is then fed into CNN and VGG-16 to retrieve the same color products from the dataset, which will reduce the browsing time of the user.

## *1.1 Organization of the Paper*

The paper is organized as follows. Section 1 introduced the significance of the work, Sect. 2 describes the techniques covered under deep learning, and Sect. 3 describes the proposed methodologies of CNN and VGG-16. Finally, Sect. 4 explains the evaluation metrics followed by the conclusion.

## **2 State of the Art**

Deep learning is a subset of machine learning, which can be thought of as a subset of machine learning. It is a field devoted to computer algorithms that learn and evolve on their own [12]. In contrast to machine learning, which employs simpler concepts, deep learning employs artificial neural networks, which are designed to mimic how

humans think and learn. This section summarizes the most popular deep learning approaches.

### **Multilayer Perceptron**

It is also known as a fully connected neural network which consists of three layers known as the input layer, one or more hidden layers(s), and output layer. As it is known as a fully connected network, each neuron is connected with every other neuron. It is extremely applicable to the tasks of classification and regression. Here, the training of data will be accomplished through backpropagation [13]. The input layer receives data and passes it on to the hidden layer. Weights are assigned to each input at random by the links between the two layers. After the weights have been multiplied separately, a bias is added to each input. To the activation function, the weighted sum is passed. The activation function can be linear or non-linear, and if the linear function is used, then it is termed a single-layer perceptron. The activation function is applied at the output layer to generate an output. The above-mentioned process is feedforward NN. Observing the loss function of output, the previous weights are adjusted and backpropagated to minimize the error.

### **Convolutional Neural Network**

It is a type of neural network that has shown to be useful in image recognition and classification applications. It is also known as the convolution net, and it is made up of three layers: convolution, pooling, and fully linked layers [9]. CNN works in a different way than normal neural networks. The height, breadth, and depth of the layers in CNNs are all determined in three dimensions. The neurons in one hidden layer only communicate with a subset of the neurons in the other layer, rather than all of them. In addition, the result is grouped along the depth dimension into a single vector. The convolution and pooling layers do feature extraction, whereas the fully connected layer performs classification.

### **Recurrent Neural Network**

A recurrent neural network (RNN) is considered as the multiple feedforward neural network, which transfers the information from one to the other. It has profound applications in speech recognition, stock market prediction, language translation, and image recognition [9]. The major drawback with the Naïve RNN is it tends to forget the old information also it suffers from the issue of vanishing gradient and exploding gradient. The problem with the feedforward network is the predictions are done based on the current input as it cannot remember the historical information. On the other hand, RNN predicts the output based on current and previously learned input. A recurrent neural network is based on the notion of preserving a layer's output and feeding it back to the input to predict the output. Hence, the nodes of a particular layer can remember the information of the past steps. The best example for RNN is Google's auto-filling words. The performance of RNN can be enhanced by extending the memory by implementing long short-term memory (LSTM) and gated recurrent unit (GRU). They have internal gates that can control information flow. These gates can learn which data in a sequence should be kept and which should be discarded.

This allows it to convey relevant information down a long chain of sequences to generate predictions.

### **Autoencoders**

Autoencoders are a specific sort of feedforward network, with an input identical to the output. It compresses the input into a lower-dimensional code and then reconstructs the result. Encoder, code, and decoder are the three parts of an autoencoder. The encoder compresses the input and generates the code, which the decoder subsequently uses to reconstruct the input. With a non-linear activation function and multiple layers, an autoencoder can learn non-linear transformations. It does not need to learn dense layers. Bottleneck refers to the layer that exists between the encoder and the decoder, i.e., the code. This is a well-designed method for determining which aspects of observed data are important and which can be ignored.

### **Restricted Boltzmann Machine**

RBMs are one of the most basic neural networks because they only have two layers: a visible layer and a hidden layer [14, 15]. In a forward pass, we feed our training data into the visible layer, and during backpropagation, we train weights and biases between them. Each hidden neuron's output is generated using an activation function like ReLU. Neurons in the same layer cannot communicate directly with one another. The connections will exist between the two layers. Since the communication is restricted, it is known as restricted Boltzmann machine.

### **Deep Reinforcement Learning**

Deep reinforcement learning (DRL) revolutionizes artificial intelligence (AI) and is a step toward designing autonomous systems to better understand the visual world [16]. A machine is continuously learning through a sequence of trials and errors, making this technology perfect for dynamic, changing situations. Although reinforcement training was available for decades, it was paired with deep learning much more recently, which provided fascinating outcomes. The deep part of reinforcement is many (deep) layers of artificial neural networks that emulate the human brain structure.

### **Generative Adversarial Networks**

GANs are deep learning algorithms that create new, training-like data instances [17]. GAN consists of two components: a generator and a discriminator. During the training data, the former generates false data and the latter can learn from the false information. GAN is used to generate realistic images, cartoons, human face photographs, and 3D objects.

### 3 Proposed Methodologies

As the work concentrates on both color and category-based image retrieval, the TF-IDF method is applied, as a preprocessing step. Then, the online scraping technique, also known as Web harvesting, is used to extract images from the product URL as the procedure focuses on color. Because there are 20,000 items in the dataset and each product can have 5 images, all of the images must be obtained from the URL. Web scraping is used to retrieve and download images from product URLs in the dataset. The extracted images are passed into convolutional neural network and VGG-16 to retrieve similar types of products. Section 4 incorporates the proof of retrieval images. Figure 1 displays the example of retrieval images. And Fig. 2 shows the product category tree.

#### 3.1 Preprocessing

Initially, the case-folding also known as the lowercasing approach is used, which turns all letters to lowercases. To eliminate the punctuation marks and non-alphabetic characters, string manipulation and regular expression procedure are used. Furthermore, a lemmatization method is applied to fetch the meaningful category for the product. TF-IDF [18] is applied, to fetch the top and least 15 frequent categories from the dataset. TF-IDF is a feature extraction technique that presents the relative importance of a term. TF is the ratio of no. of times term ' $t$ ' appears in the document to the total number of terms whereas IDF is the ratio of the number of documents to the number of documents that has a term ' $t$ ' which will compute the logarithmic further. By applying the TF-IDF, the most frequent and least frequent product categories are identified. Only the top-15 frequent categories are considered for further process. The prediction of categories can be easily identified through the above-mentioned approach.



**Fig. 1** Example of retrieval images

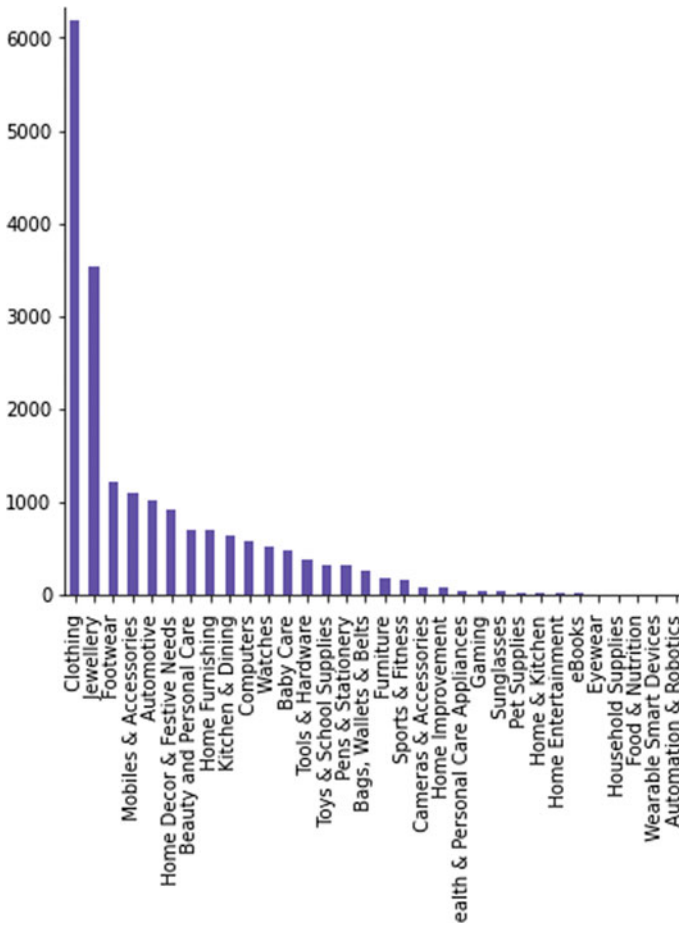


Fig. 2 Product category tree

### 3.2 Color Identification

The color feature is used to find the corresponding product based on a similar intensity of pixels. To extract color information, color histogram techniques, color coherence vectors, and color moments are often utilized. This paper utilizes the color histogram to identify the color of the product. A color histogram [19] is a graph that depicts the distribution of color composition in an image. In addition, the color histogram bar shows the number of pixels in each type of color in an image. A histogram’s data are obtained by counting each color in the image. To identify the similarity matching between the images, the Euclidean distance is computed. The formula to compute the Euclidean distance is given below in Eq. (1).

$$D = ((x_2 - x_1)^2 + (y_2 - y_1)^2)^{1/2} \tag{1}$$

The color of the retrieved images will be determined this way. Finally, the color identified through the color histogram will be passed into the DL algorithms, which will predict the matching image for the selected product.

### 3.3 Application of DL Techniques

#### 3.3.1 Convolutional Neural Network

The reason behind the choice of CNN is, when compared to other classification methods, the amount of pre-processing required by a ConvNet is significantly less. It is also known as convolution nets, which have three types of layers known as convolution layer, pooling layer, and fully connected layer. When compared to regular neural networks, CNNs work differently. The architecture of CNN for the proposed work is depicted in Fig. 3.

The image gets differed from one another in their special structure. CNN could be used to derive a higher-level representation of image content. In addition to the color identification product in the E-commerce sites, various applications can be listed for the color identification process. For instance, in the agricultural industry, to determine the color of fruits [20] that are riped completely and to discard it, to identify the color of the light under different conditions, detecting and identifying the color of the traffic signal to maintain road safety. As the implemented work is incorporated in the images taken from the e-commerce site, this method can be useful to recommend similar color products [12] by observing the user’s behavior.

In CNN, the overhead to preprocessing will be lesser. The basic elements of CNN are the convolutional layer, pooling layer, fully connected layer. The convolutional layer is responsible for feature extraction. The term ‘convolution’ is a mathematical

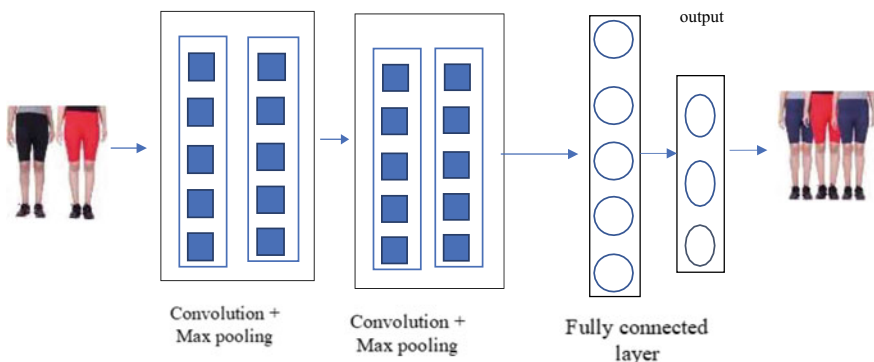


Fig. 3 Architecture of CNN



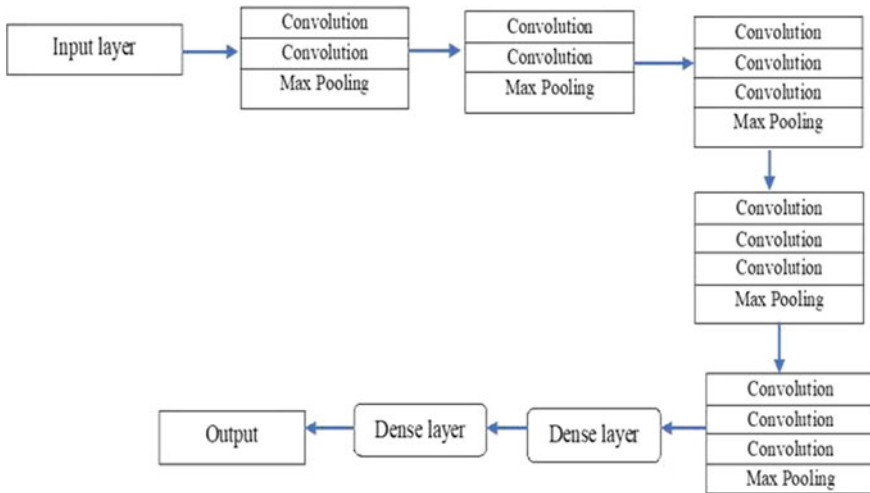
function that combines two functions and produces a third function using the merging technique. The images retrieved from the e-commerce sites are represented in terms of the pixel matrix. The input to the convolutional layer is the height, width, and number of channels in the image. As we have taken the RGB image, the number of the channel will be 3.

As the CNN will not consider the whole image of pixels, only the part of the pixels is supplied as input. For instance, instead of taking the whole pixel matrix, a small subset of matrix values will be taken. For instance, from the whole matrix, either  $3 \times 3$  or  $5 \times 5$  matrix values and the extracted feature map will be applied with the filter which is of the same size. The number of features chosen by the CNN is directly proportional to the training time. If the number of features is high, then the time taken for training will also be longer. Then, the process of activation function begins, here the 'ReLU' function is used, which will replace the non-negative value with zero. The pooling layer works similar to convolution, and there are various types of pooling available known as max pooling, min pooling, and average pooling. It slides the value obtained through convolution by considering the mentioned value. For instance, the max pooling is applied here, which will stride the max. Value from the pixel groups. The obtained values are then randomly passed to the fully connected layer. The final layer has a 'Softmax' function which will predict the recommendation for the query image by considering the category and color.

### 3.3.2 VGG-16

VGG-16 is a type of CNN model known as a very deep convolutional neural network. VGG-16 has a convolutional layer which is  $3 \times 3$  size, and the max pooling layer has  $2 \times 2$  size followed by the fully connected layers at the end. Hence, '16' denotes the number of layers that have weights [20]. The formation of convolution and max pool layer is consistent with the whole architecture. Figure 4 shows the architecture of VGG-16.

In 2014, Simonyan and Zisserman [21] presented VGG-16 architecture as a very deep convolutional network for large scale image recognition. A sequence of convolutional layers is followed by a max pooling layer in each VGG block. All convolutional layers have the same kernel size ( $3 \times 3$ ). After each block, a max pooling of size  $2 \times 2$  with strides of 2 is used to halve the resolution. There are two completely linked hidden layers and one fully connected output layer in each VGG model. The dimensions of the input image change to  $224 \times 224 \times 64$  when it is passed through the first and second convolutional layers. The output is then sent to the max pooling layer with a stride of two.



**Fig. 4** Architecture of VGG-16

## 4 Performance and Evaluation

### 4.1 Datasets

The evaluation dataset is a pre-crawled dataset derived from data extracted from Flipkart.com, a leading e-commerce site. Product URL, product name, product category tree, pid, retail-price, reduced price, image, is FK advantage product, product description, product rating, overall rating, brand, and product specification are among the fields in this collection [22]. Figure 5 shows the process of preprocessing such as stopwords removal, lemmatization, etc. Figures 6, 7 describe the top and least 15 categories of items. Figure 8 depicts the color histogram of the query image, and finally, Fig. 9 shows the proof of image retrieval based on color and category.

### 4.2 Precision, Recall, F-Measure, and Accuracy

To evaluate the performance of CNN and VGG-16, the performance metrics [23] of precision, recall, F-measure, and accuracy are applied. Precision defines, ‘how many selected items are relevant’, and recall defines, ‘how many relevant items are selected’. Accuracy defines the correctly predicted data to the total observations. The formula to calculate precision, recall, F-measure is mentioned below in Eqs. (2), (3), (4), and (5). Where TP, TN, FP, and FN denote true positive, true negative, false positive, and false negative, respectively.

```

[94] punc = str.maketrans('', '', string.punctuation)
#apply the empty mapping table to each element of the series where x is the review for one document.
df['Sub_category'] = df['Sub_category'].apply(lambda x : ''.join(word.translate(punc) for word in x.split()))
print('Punctuation Remove: ' + str(df['Sub_category'][2]))
print('\n')
#list of stop words
stop = stopwords.words('english')
#removing the stop words
df['Sub_category'] = df['Sub_category'].apply(lambda x : ''.join(word for word in x.split() if word not in stop))
print('Stopwords Remove: ' + str(df['Sub_category'][2]))
print('\n')

#lemmatize words to reduce them to their root form. Note: added the pos = 'v' to reduce the incoming word to verb root
lem = WordNetLemmatizer()
df['Sub_category'] = df['Sub_category'].apply(lambda x : ''.join(lem.lemmatize(word, pos = 'v') for word in x.split()))
print('lemmatized Text: ' + str(df['Sub_category'][2]))

Original Text: Clothing >> Womens Clothing >> Formal Wear >> Waistcoats >> KAJCI Waistcoats >> KAJCI Embroidered Womens
Punctuation Remove: Clothing Womens Clothing Formal Wear Waistcoats KAJCI Waistcoats KAJCI Embroidered Womens Waistco
Stopwords Remove: Clothing Womens Clothing Formal Wear Waistcoats KAJCI Waistcoats KAJCI Embroidered Womens Waistcoat
lemmatized Text: Clothing Womens Clothing Formal Wear Waistcoats KAJCI Waistcoats KAJCI Embroidered Womens Waistcoat
    
```

Fig. 5 Preprocessing the dataset

```

[27] unique_labels = set(df['Possible_category'])
print('Number of unique labels:', len(unique_labels))

Number of unique labels: 995

from collections import Counter
Sub_category_count = Counter(df['Possible_category']).most_common()

print('15 most frequent Category:')
print(*Sub_category_count[:15], sep = "\n")
print('\n 15 least frequent Category:')
print(*Sub_category_count[-15:], sep = "\n")

15 most frequent Category:
(' Rings ', 207)
(' Necklaces & Chains , Necklaces ', 118)
(' Bed Linen , Blankets, Quilts & Dohars ', 100)
(' Tools , Gardening Tools , Plant Containers & Sets , Plant Container Sets ', 58)
(' Curtains & Accessories , Curtains ', 27)
(' Decorative Lighting & Lamps , Series Lights , Digilight Series Lights ', 27)
(' Wall Decor & Clocks , Clocks , Wall Clocks , Blacksmith Wall Clocks ', 20)
(' Earrings ', 20)
(' Womens Footwear , Casual Shoes , Boots ', 15)
(' Mens Clothing , T-Shirts , Dvlyon T-Shirts ', 15)
(' Coffee Mugs , Rockmantra Coffee Mugs ', 14)
(' Womens Clothing , Ethnic Wear , Kurtas & Kurtis , Kurtas , Parijat Kurtas ', 14)
(' Jewellery Sets ', 13)
    
```

Fig. 6 Retrieving the top 15 frequent category

$$\text{Precision}(P) = TP/TP + FP \tag{2}$$

$$\text{Recall}(R) = TP/TP + FN \tag{3}$$

```
+ Code + Text
(* Beauty and Personal Care ', 87)
(* Footwear ', 86)
(* Tools & Hardware ', 70)
(* Mobiles & Accessories ', 66)
(* Computers ', 65)
(* Kitchen & Dining ', 56)
(* Baby Care ', 47)
(* Automotive ', 45)
(* Matches ', 31)
(* Bags, Wallets & Belts ', 27)
(* Sports & Fitness ', 26)

15 Least frequent Category:
(* Baby Care ', 47)
(* Automotive ', 45)
(* Matches ', 31)
(* Bags, Wallets & Belts ', 27)
(* Sports & Fitness ', 26)
(* Gaming ', 24)
(* Toys & School Supplies ', 16)
(* Pens & Stationery ', 15)
(* Home Improvement ', 15)
(* Furniture ', 15)
(* Home Entertainment ', 5)
(* Home & Kitchen ', 4)
(* Sunglasses ', 3)
(* Health & Personal Care Appliances ', 3)
(* Pet Supplies ', 1)
Activate Windows
Go to PC settings to activate Windows.
```

Fig. 7 Retrieving the least 15 frequent category

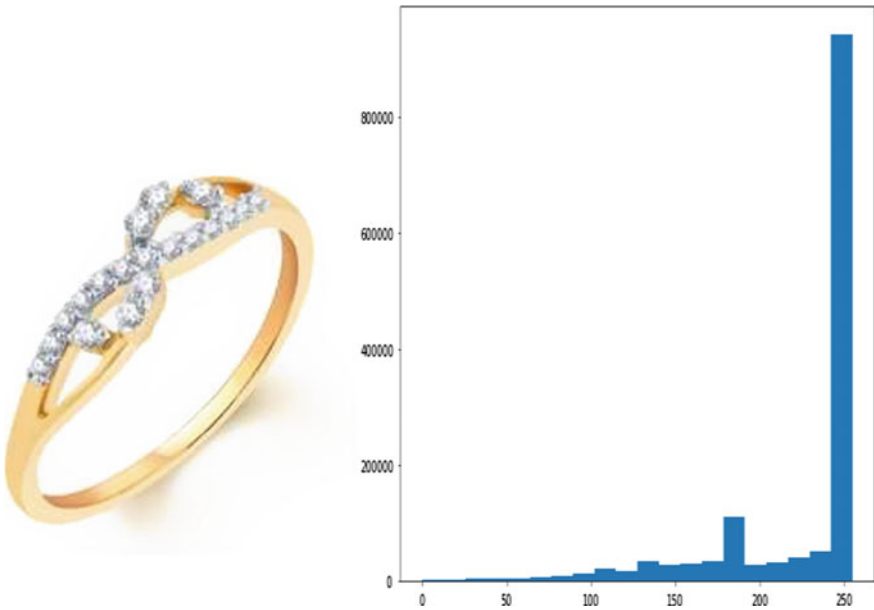


Fig. 8 Color histogram of the query image

$$F\text{-Measure} = 2 * (P * R)/(P + R) \tag{4}$$

$$Accuracy = TP + TN/TP + FP + FN + TN \tag{5}$$



**Fig. 9** Category and color-based retrieval

**Table 1** Comparison of performance

Metrics/techniques	CNN	VGG-16
Precision	73.52	90.47
Recall	87.71	95
F-measure	0.8	0.92
Accuracy	85.32	91.25

Table 1 shows the obtained results by applying CNN and VGG-16. And Fig. 9 shows the similar product retrieved by applying the above technique.

From the obtained results, it can be concluded that the results of VGG-16 outperform the CNN.

## 5 Conclusion

This paper adopts the strategy of convolutional neural network and VGG-16 to fetch the color-based products. Also, 20,000 images are trained to predict the products based on color. Initially, the TF-IDF is applied to identify the category of the product. Then, the Web scraping technique retrieves all the images from the product URL, and it is fed into a color histogram, to identify the color of the product. From the color histogram, the color of the product is easily identified and it is passed into CNN and VGG-16 to fetch the matching images for the chosen product. The experimental results show a good prediction for the implemented models. The evaluation metrics of precision, recall, F-measure, and accuracy demonstrate that our system performs well with the trained model, and VGG-16 performs better with CNN. This paper concentrates on image retrieval in the domain of apparel, and further, it can be enhanced to various products available in the dataset. This image retrieval can be implemented on e-commerce sites to reduce the browsing time of the user.

## References

1. Guan C, Qin S, Ling W, Ding G (2016) Apparel recommendation system evolution: an empirical review. *Int J Clothing Sci Technol* 28(6):854–879
2. Tewari AS (2020) Generating items recommendations by fusing content and user-item based collaborative filtering. *Procedia Comput Sci* 167:1934–1940
3. Pazzani M, Billsus D (2007) Content-based recommendation systems. In: Brusilovsky P, Kobsa A, Nejdl W (eds) *The Adaptive web*, lecture notes in computer science, vol 4321. Springer Berlin Heidelberg, pp 325–341
4. Chu W-T, Tsai Y-L (2017) A hybrid recommendation system considering visual information for predicting favorite restaurants. *WWWJ* 1–19
5. Wang Y-F, Chuang Y-L, Hsu M-H, Keh H-C (2004) A personalized recommender system for the cosmetic business. *Expert Syst Appl* 26(3):427–434
6. Blake MB (2017) Two decades of recommender systems at Amazon
7. McAuley J, Targett C, Shi Q, van den Hengel A (2015) Image-based recommendations on styles and substitutes. In: *Proceedings 38th international ACM SIGIR conference research developmental information retrieval*. pp 43–52
8. Goodfellow I, Bengio Y, Courville A (2016) *Deep learning*. MIT Press. <http://www.deeplearningbook.org>
9. Zhang S, Yao L, Sun A, Tay Y (2019) Deep learning-based recommender system: a survey and new perspectives. *ACM Comput Surv* 52(1):38, 5. <https://doi.org/10.1145/3285029>
10. Sabahi F, Ahmad MO, Swamy MNS (2016) An unsupervised learning-based method for content-based image retrieval using hopfield neural network. 2016 2nd international conference of signal processing and intelligent systems (ICSPIS). pp 1–5. <https://doi.org/10.1109/ICSPIS.2016.7869882>
11. Wei W, Wang Y (2019) Color image retrieval based on quaternion and deep features. *IEEE Access* 7:126430–126438
12. Deng L, Yu D (2014) Deep learning: methods and applications. *Found Trends<sup>®</sup> Sig Process* 7(3–4):197–387
13. Zhang P, Jia Y, Gao J, Song W, Leung H (2020) Short-term rainfall forecasting using multi-layer perceptron. *IEEE Trans Big Data* 6(1):93–106. <https://doi.org/10.1109/TBDATA.2018.2871151>
14. Salakhutdinov R, Mnih A, Hinton G (2007) Restricted Boltzmann machines for collaborative filtering. In: *Proceedings of the 24th international conference on Machine learning, ICML, association for computing machinery*. New York, NY, USA, pp 791–798
15. Liu X, Ouyang Y, Rong W, Xiong Z (2015) Item category aware conditional restricted Boltzmann machine based recommendation. In: *International conference on neural information processing*. Springer, pp 609–616
16. Arulkumaran K, Deisenroth MP, Brundage M, Bharath AA (Nov 2017) Deep reinforcement learning: a brief survey. *IEEE Sig Proc Mag* 34(6):26–38. <https://doi.org/10.1109/MSP.2017.2743240>
17. Cai X, Han J, Yang L (2018) Generative adversarial network based heterogeneous bibliographic network representation for personalized citation recommendation. In: *AAAI*
18. JingL-P, Huang H-K, Shi H-B (2002) Improved feature selection approach TFIDF in text mining. In: *Proceedings of the international conference on machine learning and cybernetics*, vol 2. pp 944–946
19. Juang C-F, Sun W-K, Chen G-C (2009) Object detection by color histogram-based fuzzy classifier with support vector learning. *Neurocomputing* 72(10–12):2464–2476
20. Yang Z, Yue J, Li Z, Zhu L (2018) Vegetable image retrieval with fine-tuning VGG model and image hash. *IFAC-PapersOnLine* 51(17):280–285

21. Simonyan K, Zisserman A (2015) Very deep convolutional networks for large-scale image recognition. CoRR [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
22. <https://www.kaggle.com/PromptCloudHQ/flipkart-products>
23. Gunawardana A, Shani G (2009) A survey of accuracy evaluation metrics of recommendation tasks. J Mach Learn Res 10:2935–2962