**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

# Recommending Items based on Image Similarity Metric

**Poonam Saini**
**Dr. Priyanka Shaktawat**

## Abstract

In an online e-commerce portal like Amazon or Flipkart which have millions of products to choose from, a problem that fast arises as to how to enhance the process of recommendations to the user with right kind of items to choose from. The existing systems are based on collaborative or content-based filtering models, which somehow generate recommendations which may not suit the user's requirements. Therefore, there is a need to strengthen the existing systems which have plenty of items, with an image-based product similarity technique to gather similar products as choices for the actual user requirement. The proposed work involves two different approaches(i) using the histogram of gradients for feature extraction from the images and then selecting the best similar choices using the K-means clustering unsupervised learning principle (ii) using the pretrained CNN model to convolute the similarity of images and give a set of five best item images. In this research work, overall, 50% accuracy has been achieved by the first approach in extracting images as per the query image posed by the user, on the other hand, the second approach has achieved 96.18% accuracy while using MobileNet based pretrained model. From this research work, it can be concluded that the image-based approach using the pretrained CNN can be used for recommending items to the users using image similarity approach. The present work also demonstrates a comparative analysis of Normalized Discounted Cumulative Gain that further assesses the recommendations made to the users.



***Keywords:***

Deep Learning;
Recommender Systems;
Image Database;
HOG;
CNN;
NDCG.

***Author correspondence:***

Poonam Saini,
Bhupal Nobles' University, Udaipur
Email: poonam.saini9@gmail.com

## 1. Introduction

With large scale improvements in internet technology, a mushroom growth of the e-commerce online portals has been observed. Almost all the manufacturers who have a product to sell in the market have adopted online-platforms as a means of reaching out to the customers and improving their sales. E-commerce has become the master ruler of any business today. E-commerce websites like Amazon, Myntra, Flipkart, Meesho etc. have redefined the term commerce and added a new meaning to it. Their web-portals are handling millions of transactions on day today basis. The role of product recommender systems cannot be denied, in fact with the advent of recommender system-based technologies, the e-commerce business has experienced unprecedented spike in attracting customers, recommending products, displaying items and comparing similar items and above all retaining customers. This transformation has come up because of the use of artificial intelligence and machine learning in all the aspects of e-commerce and business. With the advancement in the computer technology and ever-increasing processor power, the process of storing the information, processing the information and the retrieving the information has become easier and faster. The use of mobile technology, where the e-commerce apps help the customers in not only recommending the items but also display the product images with sufficient product information and price tagging and smoothens the online business transactions. The mobile devices in comparison with desktops and laptops have equally powerful processors, which have given a momentum to the online industry.

**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

The recommender systems are machine learning algorithms whose basic task is to recommend products to the customers. The recommender systems improve the overall product selection and provides the users with a set of best choices in a short period of time without any hassles, thereby improving the overall decision-making process. In order to accomplish this task, the recommender system needs to have some prior information, in terms of users' past buying behavior, sometimes the social connectivity of the user etc. The recommendation systems just revolve around the collection of useful information, learning from the past and present inputs to the system and then producing the information nuggets. The recommender systems serve the customers as per their affective states during the buying process. The recommender systems collect the information or get the feedback from the customers explicitly, implicitly or in a hybrid mode. The explicit mode receives the information directly from the customers about the liking for items. The implicit feedback is obtained by observing the user while he or she is browsing the content on the web portal. It means their behavior, their clicks on the information pages are observed at the backend. The third technique is a combination of the implicit and explicit techniques and also called as hybrid technique. Therefore, the ultimate goal of any recommender system algorithm is in providing ranked and classified business information to the end users.
The ultimate purpose of the recommender system is to enhance the sales of the e-commerce companies and accounting to the operational and technical goals. According to [1], the operational and technical goals of a recommender system can be divided as follows as shown in figure 1.


Figure 1. A Recommender System overview

The sole purpose of any recommender system is to see that the old recommendations are the ones that are least repeated, else this may lead to irritated environment, thereby generating dissatisfaction on e-commerce portal, and thereby decrease in the overall sales process or even the user might leave the portal for the product elsewhere. There should be an element of surprise for the user, thereby making him filled with awe and surprise. The diversity of items helps in ensuring that the user who needs choices does not get the repetition of the items and most importantly the recommended items should be relevant to the query of the end user.
Theoretically, there are three basic types of recommender systems: (i) Collaborative filtering (ii) Content based recommender systems (iii) Hybrid Recommender Systems
The concept of collaborative filtering involves the supplying of information regarding the products to the end user according to the user's taste and the preferences. The system builds a user profile by amalgamating the user's similarities with other users. The content-based recommender systems revolve around the content or the metadata related to the items or customers. For the items, it could be the item code, item name, its description, item images, information related to the item group or sub-groups or any other technical specification etc. This technique recommends the items by assessing the most similar or related item or those items that have certain degree of similarity of content. The hybrid recommender systems leverage the combined effects of collaborative and content-based recommender systems. The present work implements the content-based recommender systems by using the item images. According to [2] the image-based recommendations have considerable impact on the customer's decision-making process. The convolutional Neural Networks (CNN) models expedite the process of extraction of useful visual information. The user sends a query to the system on an online e-commerce platform, for expressing his desire in buying a product from the store. The recommendation system in turn recommends products according to features extracted from product images. The basic task involves recommending item images that are close to the query image.

This image-based recommender system recommends items on the basis of recent shopping history or interaction with the e-commerce platform. The present work involves the design of a process to recommend most suitable items to the users on the basis of the current selection and choice of the product. The contribution of the present work involves:

- Calculation of HOG(Histogram of Oriented Gradients) features for the image dataset and then classify on the basis of K-means clustering which is an unsupervised learning algorithm.
- Using a five pretrained CNN model (VGG16, ResNet50, VGG19, Xception and MobileNet) in extracting the important image feature and then recommending the most relevant items and then calculating the accuracy of recommendations thus made.
- Calculating the Normalized Discounted Cumulative Gain (NDCG) to further assess the recommendations made.

This work has been divided into different sections where Section-2 describes existing literature available using an image-based approach in Recommender Systems, Section-3 expresses the proposed methodology, Section-4 describes the result and discussion portion and in Section-5 conclusion has been drawn and suggested future studies have been expressed.

## 2. Review of Literature

Images have always been a major source of attraction for the customers, be it for the clothing, grocery, or electronics items [2]. The researchers have worked with the design and implementation of the recommendation systems using images. This work involves the use of perceptual retrieval, Gaussian mixture models, Markov Chains, Texture Agnostic Retrieval methods to assess the overall recommendations thus generated.

In [3] a detailed study has been carried out in classifying 1000 different images classes using Alexnet and [4] describes the details about another form of CNN architecture: VGG that was the most powerful classifier in Imagenet competition in the year 2014. Both [3, 4] describe a detailed classification process for the images and recommend on that basis.

Most of the research work that has been carried out in the past relates more to the identification of the class that an object or an item belongs to, but there can be subclasses as well, that an item may be a part of it. There can be variations within the class [4, 5]. The visual features extracted from the product images have certain characteristic features that could be utilized for the matching purpose with that of the other images in the database.

The research work of [6] has utilized the image features in creating a visual content-enhanced POI recommender system (VPOI). The researchers [7] have designed a search engine for online shopping using the Amazon dataset and two CNN models, including VGG and AlexNet which are a set of pre-trained models. They were able to improve the classification accuracy and have used Jaccard similarity to calculate the similarity score.

The researchers [8] have designed an image-based recommendation system that uses a fashion dataset to train a CNN model to solve image classification process. The researchers designed a visually aware feature extractor to feed the ranking system. This ranking system provided the recommendations from most appropriate recommendation to the least appropriate ones.

The researchers [9] have designed an aesthetic-based clothing recommendation method based on cross product of matrix and tensor factorization methods. The image features and the aesthetic features were extracted and utilized in the recommendation process.

The researchers [10] developed a restaurant recommendation method based on image data effectiveness, including images of food and restaurants. They also used the CNN model to extract visual features and text representation of input data

## 3. Proposed Methdology

The basic work of a recommender system is to recommend the items to the users on the basis of the information received from the system and in present research work, it is the product query image that has been enquired or posted by the end user. The first and the foremost task is to identify the main category to which the query-item-image belongs to and then search the main database and display the most similar or relevant items that have been ranked in an order, infront of the user. To execute the entire process, this present work has utilized (i) Five pretrained the convolutional neural networks (CNN) models (ii) the HOG(Histogram of Oriented gradients) based image processing technique.

**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

3.1 Convolutional Neural Network Architecture

Convolutional Neural Networks (CNN) are the feedforward neural networks which utilize a varying set of hidden layers, fully connected layers and SoftMax layer. The size of the input and the output data are fixed. The input data is in this case are the images and the output is the categorization of the input with its confidence level. Figure 2 shows that there are hidden layers and a classification unit. The hidden layers have several sets of a combination of convolutional layers and a ReLU unit with pooling layers. The classification unit has a SoftMax layer for final classification. The entire CNN unit takes in input images of specific sizes as shown in figure 2. The CNN model was designed by Kunihiko Fukushima, a researcher at the NHK Broadcasting Science Research Laboratories in Kinuta, Setagaya, Tokyo, Japan [10]. The team of Yan LeChun further improved the CNN model and named the new CNN model as LetNet, which was basically utilized for the number and handwriting recognition tasks.

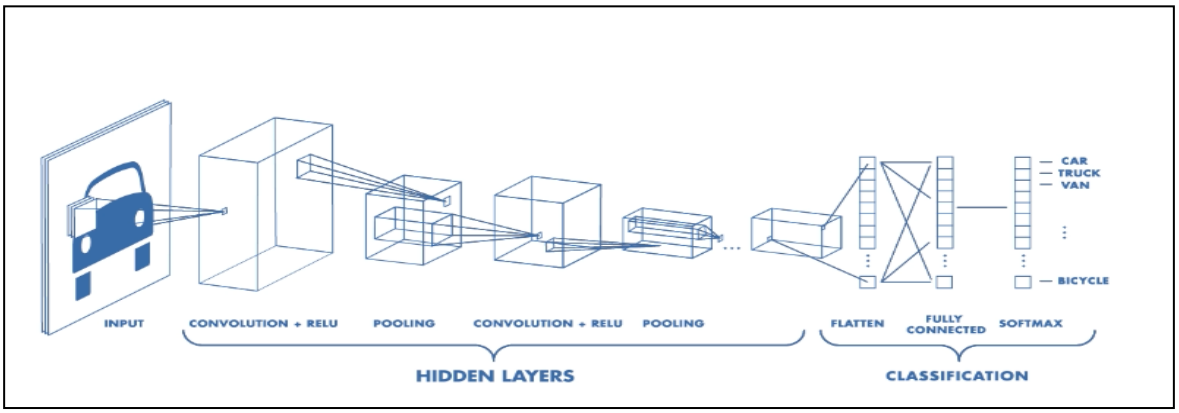


Figure 2. An architecture of Convolutional Neural Networks[11]

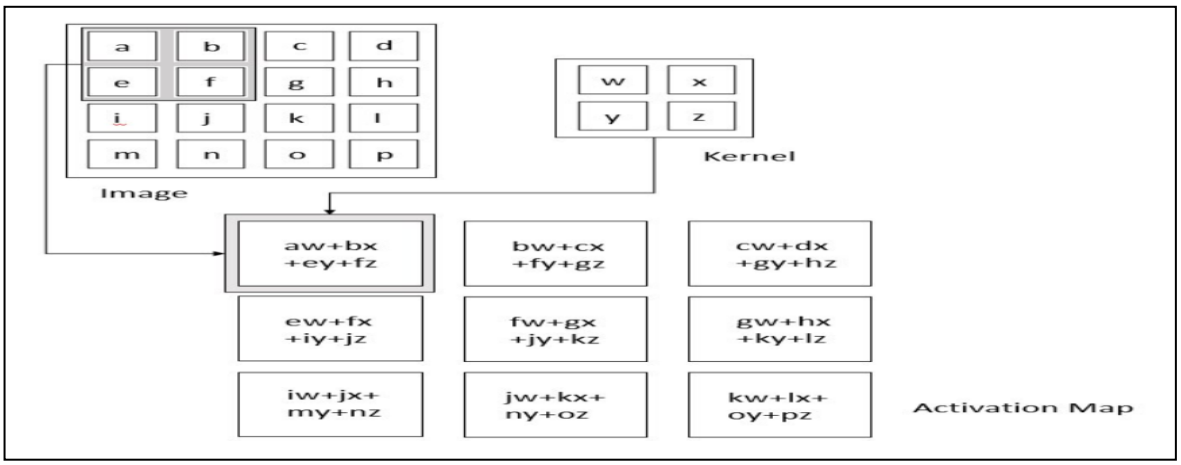


*Figure 3. Convolution Operation [12]*

The Figure 3, shows the convolutional operation between an image and the kernel. The input image is 4 X 4 matrix and the kernel or the filter is 2 X 2 matrix. Since the image is 4X4 and the filter is 2 X 2, then the output after convolution shall be 3X3 matrix as shown in the figure 3.

If we have an activation map of size $P$ $X$ PX $D$, a pooling kernel of spatial size $F$, and stride $S$, then the size of output volume can be determined by (1). Since in figure 3, P=4, and F=2, if S=1, then $P_{out}$ shall be 3. Therefore, $P_{out}$=3 X 3 matrix.

$$P_{out} = \left(\frac{P-F}{S}\right) + 1 \quad (1)$$

A fully connected Layer as shown in the figure 2, has a connection with all the layers of the entire network. The fully connected layer helps in mapping the input and the output. The ReLU or the Rectified Linear Unit is a popular activation function as it can converge six times faster than other activation functions like tanh and

sigmoid. ReLU seems to be a linear function. But in fact, it is a non-linear function, and it is required so as to pick up and learn complex relationships from the training data. The derivative of an activation function is required when updating the weights during the backpropagation of the error. The slope of ReLU is 1 for positive values and 0 for negative values. It becomes non-differentiable when the input value is zero, The ReLUis shown in (2):

$$f(k) = \max(0, k) \quad (2)$$

3.2 Models Used

The present work has used VGG16, RESNET50, VGG19, Xception and MobileNET based pretrained CNN models for the recommendation process which have been described briefly as follows:

- VGG16: The VGG net or VGG model is pretrained CNN model with 16 layers. It was designed by K. Simonyan and A. Zisserman from Oxford University who proposed this model and published it in [13, 20]. The VGG16 model can achieve a maximum test accuracy of 92.7%, on a dataset containing more than 14 million training images across 1000 object classes. The VGG16 model has sixteen layers having weights, there are thirteen convolutional layers, five Max Pooling layers, three dense layers. The total comes to twenty-one, but only the sixteen weight layers have weights to attain learnable parameters. VGG16 model tales in 224X224X3 color image. VGG16 model has tuneable hyperparameters with 3X3 filters with stride 1 and always used the same padding and maxpool layer of 2 X 2 filter of stride 2. The Conv-1 layer has 64 number of filters, Conv-2 has 128 filters, Conv-3 has 256 filters, Conv 4 and Conv 5 has 512 filters. There are three fully connected layers (FC) which have a huge stack of convolutional layers. The first two FCs have 4096 channels and the third layer is used for classification of 1000 classes having one thousand channels. The final layer is the soft-max layer.
- ResNet50: A Resnet50 architecture consists of 50 layers and the details are as follows: The first layer is made up with a kernel size of 7 * 7 and 64 different kernels and all have a stride of size 2. The second layer has a max pooling layer with a stride size of 2. The next structure has a set of 9 layers (next 3 layers are repeated 3 times) . The next convolutional layer consists of 1*1, 64 kernels, subsequently, there is a convolutional layer of 3*3 with 64 kernels. The subsequent layers consist of 1*1 with a 256-kernel structure. This way there are 9 layers. The next stage consists of 12 layers (next four layers are repeated 3 times). The first layer of this section has 1*1, 128 kernels, next with 3*3, with 128 kernels and 1*1 with 512 kernels, thereby creating 12 layers with a repetition. Again, there is a set of 18 layers in a pair of 6 layers. This structure has a kernel of 1 * 1,256 and two more kernels with 3 * 3,256 and 1 * 1,1024 and this is repeated 6 time giving us a total of **18 layers**. The next structure consists of 1 * 1,512 kernel with two more of 3 * 3,512 and 1 * 1,2048 and this was repeated 3 times giving us a total of **9 layers**. In the end an average pool layer is there with a fully connected layer containing 1000 nodes, finally it is followed by a SoftMax layer. Thus, the entire structure gives us a 1 + 9 + 12 + 18 + 9 + 1 = 50 layers.[14].
- VGG19: VGG19 model has 19 layers and was trained with (224 X 224) RGB image set. It has used kernels of (3 X 3) size with a stride size of 1 pixel. The max pooling was performed over a 2 X2-pixel windows with stride 2. Then follows the Rectified linear unit(ReLu) to introduce non-linearity to make the model classify better and to improve computational time as the previous models used tanh or sigmoid functions. It has used three fully connected layers and a SoftMax layer in the last [15].
- Xception: Xception is a convolutional neural network architecture based entirely on depth wise separable convolution layers. The Xception architecture is a linear stack of depth wise separable convolution layers with residual connections. This makes the architecture very easy to define and modify; it takes only 30 to 40 lines of code using a high level library such as Keras or TensorFlow [16].
- MobileNet: A MobileNetV2 has **inverted residual structure**and its n**on-linearities in narrow layers has been removed.** It is finding application in the areas of for feature extraction, object detection and semantic segmentation. In MobileNetV2, there are two types of blocks,one is residual block with stride of 1 and another one is block with stride of 2 for downsizing. There are 3 layers for both types of blocks[17].

3.3 Image Similarity Metrics

There are a variety of mathematical techniques that are available to compute the similarity between the two images of interest.

- Cosine Similarity: A Cosine similarity is a metric that is used for computing the similarity of two items. Mathematically, it measures the cosine of the angle between two vectors projected in a multi-dimensional

**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

space. The output value ranges from **0–1**. A *'0' means no similarity, whereas '1' means that both the items are 100% similar.*

$$cos\ \theta = \left(\frac{P.Q}{||P||x||Q||}\right) = \frac{\sum_{i=1}^{n} P_i x Q_i}{\left(\sqrt{\sum_{i-1}^{n}(P_i)^2}\right)\left(\sqrt{\sum_{i-1}^{n}(Q_i)^2}\right)} (3)$$

Equation (3) shows the formula for the computation of cosine similarity between two items.
In (3) 'P' and 'Q' are the vectors, the dot product of vectors P and Q is (P.Q), where ||P|| denotes the length of vector P, while ||Q|| denotes the length of vector Q. ||P||x||Q||| is the cross product between the two vectors P and Q.

- Euclidean Distance: The Euclidean distance between the two points $P(p_1,p_2)$,$Q(q_1,q_2)$ on the cartesian coordinate system is given by (4):

$$d(P,Q) = \sqrt{(q_1 - p_i)^2 + (q_2 - p_2)^2} (4)$$

3.4 NDCG (Normalized Discounted Cumulative Gain)
The entire base lifeline of any recommender system is its ranking model which needs regular updating. NDCG is the concept used for ranking the quality. The performance of the recommendation engine is evaluated on the basis of NDCG.

$$NDCG = \left(\frac{rel\,(Items\ \ returned\ \ by\ the\ searc\ h\ engine}{rel\,(Items\ \ returned\ \ by\ the\ ideal\ -searc\ h\ engine\ )}\right) (5)$$

The NDCG measure is used by the researchers to evaluate the approach in this case. The NDCG is one of the criteria used to rank search engine results. Meanwhile, in (6), the $DCG_p$ average is defined where $rel_p$ represents the relevant items int the corpus at position p.

$$DCG_p = \sum_{i=1}^{p}\left(\frac{rel_i}{log_2(i+1)}\right) (6)$$

$$iDCG_p = \sum_{i=1}^{|rel_p|}\left(\frac{rel_i}{log_2(i+1)}\right) (7)$$

While $NDCG_p$ is the standard form of $DCG_p$, it can be defined using (8), where $DCG_p$ is the recommendation sequence's DCG value and $iDCG_p$ is the ideal sequence's DCG value. The formula 6 and 7 represent the required equations.

$$NDCG_p = \left(\frac{DCG_p}{iDCG_p}\right) (8)$$

3.5 HOG Features of the Images
HOG (Histogram of Oriented Gradients) features have been widely used in image processing tasks like object detection and recognition tasks, pedestrian detection, face detection, and vehicle detection etc[18]. HOGs are relatively simple to compute and are effective at capturing the shape and structure of objects in images. In the present work, HOG feature extraction has been carried out using scikit-image and OpenCV libraries.HOG is a feature descriptor commonly used in computer vision and image processing. It is a technique i.e., used for object detection and image classification that involves computing and analysing the distribution of gradient orientations in an image. In simple terms, HOG features describe the shape or edge structure of an object in an image by counting the occurrences of different orientations of image gradients in localized portions of the image. The HOG feature descriptor calculation involves a five-step procedure. The process involves collecting the image and performing the image preprocessing steps including grayscale conversion and contrast equalization. The next step involves the calculation of image gradient, the third step involves the gradient binning and the fourth one involves the block normalization process and finally the normalized histogram is converted into a single feature vector. The steps are as shown through figure4. The Figure 5 shows the complete procedure followed. Once the HOG features have been extracted, they are subjected to the process of K-means clustering technique to identify five closest or most similar images available in the dataset.

**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A.

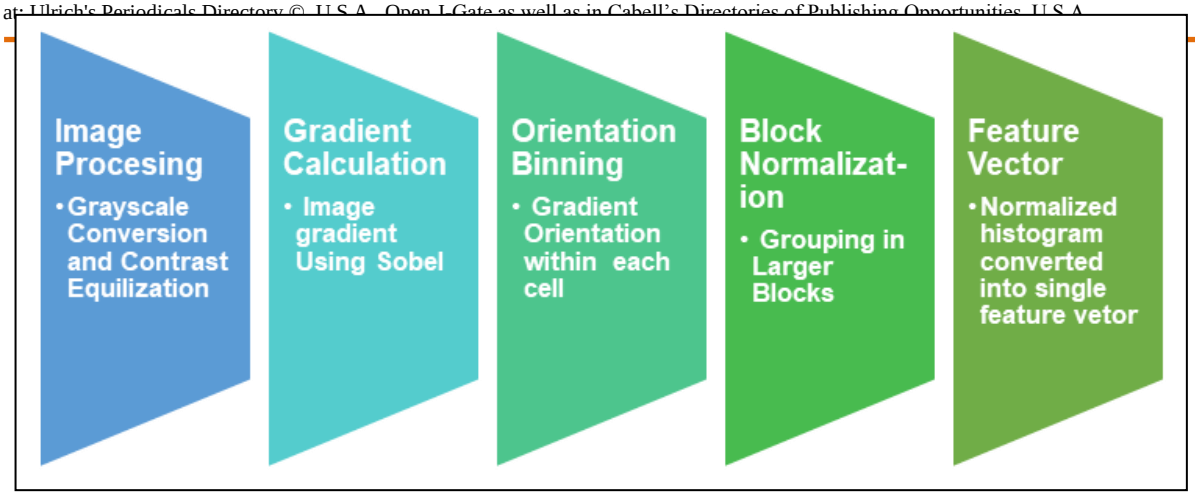


Figure 4. HOG feature descriptor calculation process

3.6 Understanding the Data

The image dataset has been collected from Kaggle [19], it has a set of 2184 images of ten different products from seven different companies. Figure 6 shows the sub-sample from the main dataset, Figure 7 shows the distribution of each item in the dataset and figure 8 shows the distribution of the items according to the company offering it. The dataset consists of images of shoes, lipsticks, handbags, nailpolish, necklace, watches, rings, bracelets, earrings and books. The Figure 7 shows the item wise distribution of objects in the dataset and the subsample in figure 6 clearly displays the type of items available in the dataset.

**4. Results and Discussion**

The entire work is based on the usage of item/product images and subsequently recommending the unique items to the users.

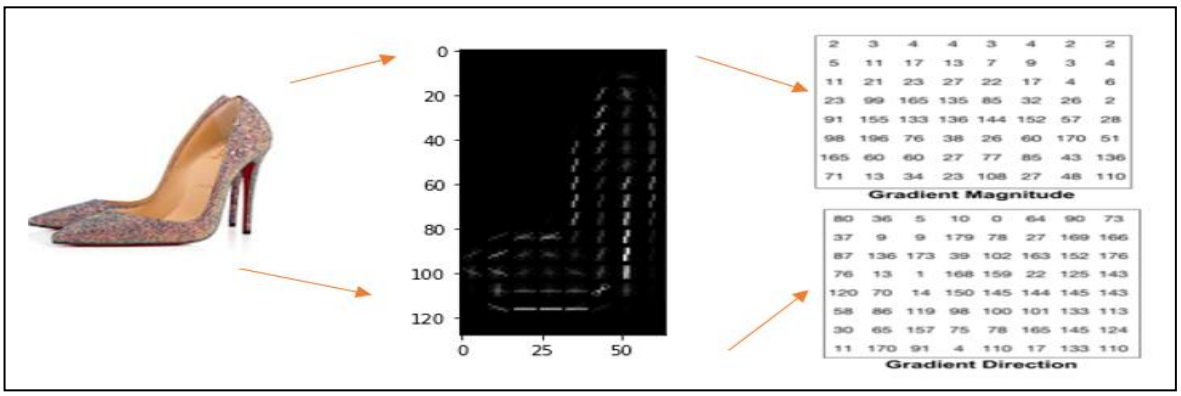


Figure 5. Histogram of Oriented Gradients

**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Figure 6. A sub-sample of the Original Dataset

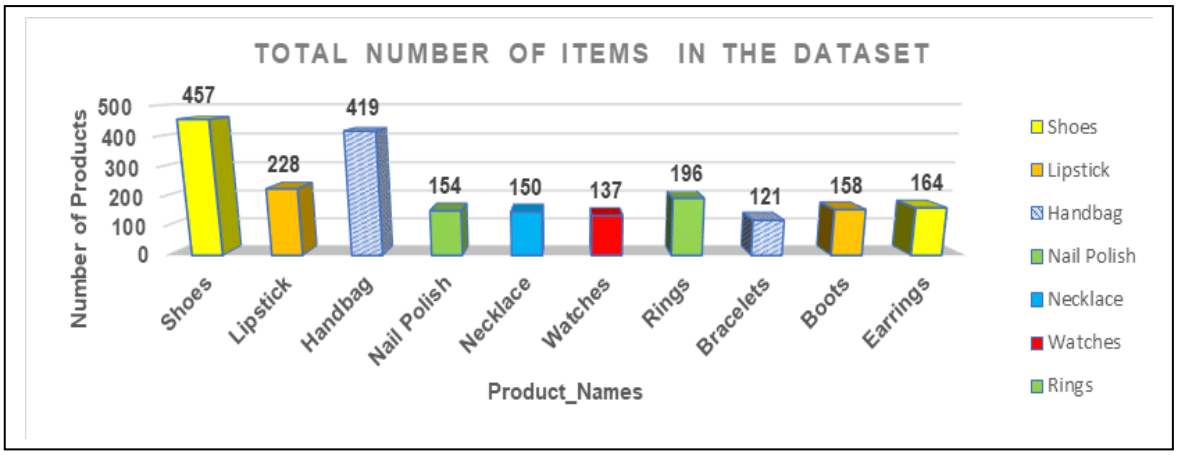


Figure 7. Product-wise distribution of items in the dataset

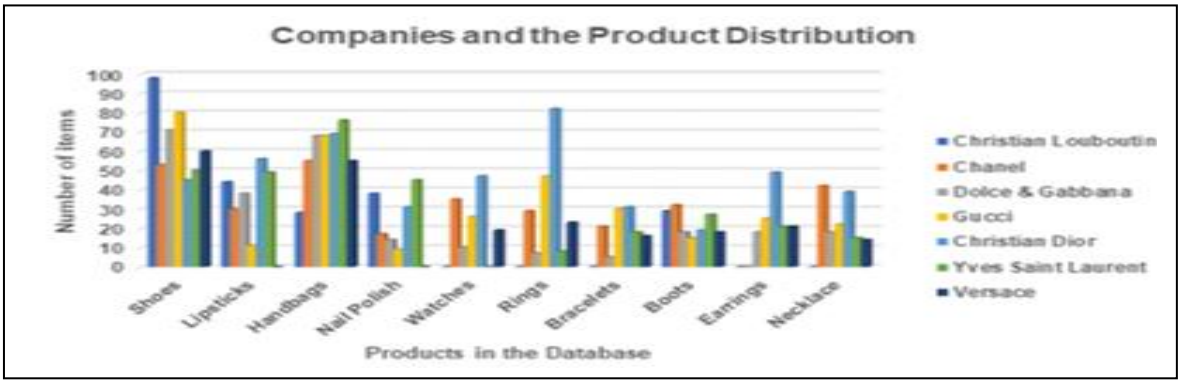


Figure 8. Distribution of items as per the company offering it.

The Figure 5, clearly demonstrates the concept of HOG(Histogram of Oriented Gradients) for one such item from the dataset. The entitre dataset of images was subjected to the proceess of caluculation of HOG parameters.The first approach adopted involves the usage of Histogram of Oriented gradietns (HOG) features for feature extraction process. The procedure involeves the calculation of HOG features for the image presented by the user to the system and then these featues were compared with the HOG features of the datset images.The comparison was carried out on the basis of the distance metric as mentioned in equation (4). The system selected five images of the data-itemsfrom the dataset with minimum value of euclidean distance. The system produces a set of five images which were were collected for repeated input sampels and resulted in the overall accuracy value of 50% using KNN algorithm. The Figure 11, clearly demonstartes the principle adopted in this research work, where an image(Original Image) of the item/product was produced in from the system and the algorithm generated a set of five similar images on the basis of the HOG features having maximum similarity. Since the overall accuracy for the procedure adopted on the basis on the basis of the HOG features being subjected to KNN algorithm generated an accuracy of 50%, which is a quite low value then another approach was used based on the CNN based five pretrained models including VGG16,VGG19,MobileNet and Xcpetion Net were tried on the existing dataset. It was observed that the overall accuracy of picking the right recommendations reached to a level of 96.18% for MobileNet based pretarined Convolutional Networks.

**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

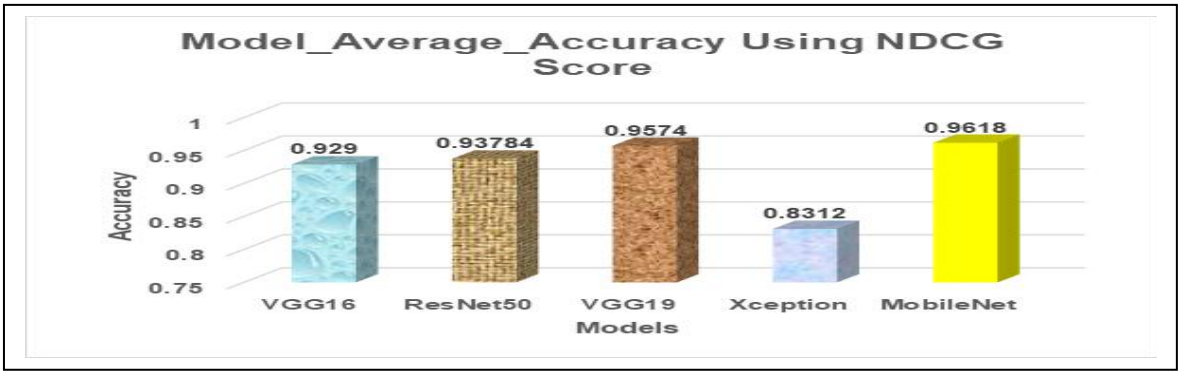


Figure 9. Results of the different Pretrained Models used.

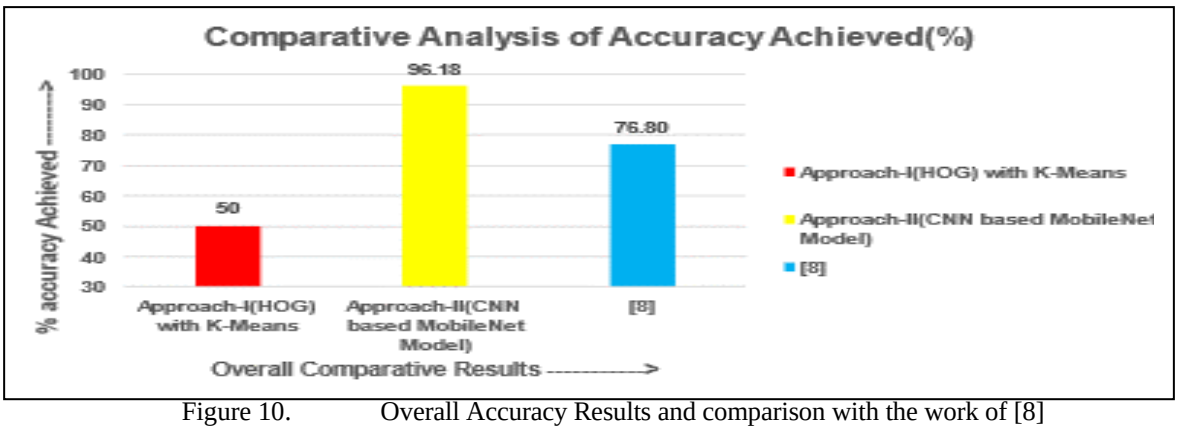


Figure 10. Overall Accuracy Results and comparison with the work of [8]

These are complex models but have been pretrained on a large dataset.If we compare the accuracy results within the set of pretrained model as shown in Figure 9, it was observed that VGG16 has an accuracy value of 92.9%,ResNet50 with 93.78%,Xception with 83.12% and finally MobileNet has an accuracy of 96.18%. All the networks have almost the same level or nearly same level of accuracy values. The main functioning of any recommender system lies in its ranking model and here in this case NDCG(Normalized Discounted Cumulative Gain) which is based on the principle of the ratio of relevance of the itemreturened by the search engine to the relevance of the items returned by the Ideal search engine.

The individual pretrained model results have been depicted through the figure 9 and figure 10 shows our comparative results. The results obtained through the pretrained model(MobileNet) fares better as compared to the HOG based feature extraction approach. The Figure 10 also shows a comparison of present research work results with [8]. The figure clearly depicts the efficacy of the pretrained models. Here Figure 11 shows the wish item image originally posted by the user and subsequent five images are the nearest recommendations made by the system using the HOG based feature extraction approach. These five recommended images have been arranged and displayed in the descending order of cosine similarity index. Similarly, the Figure 12 shows a set of multiple images (twenty in our case, the top first row of images of the figure 12) selected in bulk and forms the part of wish-list for a user. The second row onward, there are five item images displayed and have also been arranged in descending order of cosine similarity index. The recommendations in the Figure 12 are the results of the pretrained model (MobileNet). The K-Means clustering results for the HOG based approach of recommendations reach an accuracy level of 50%, whereas the accuracy attained by pretrained model-MobileNet is 96.18%. The MobileNet based CNN model outperforms and as depicted in the figure 9.

**International Journal of Management, IT & Engineering**
Vol. 13 Issue 09, September 2023
ISSN: 2249-0558 Impact Factor: 7.119
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

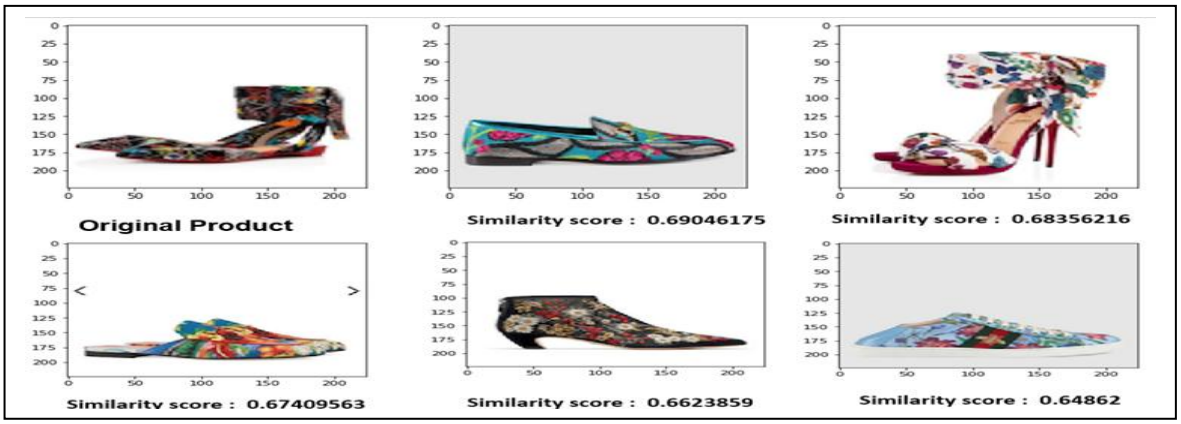


Figure 11. Using HOG features and then recommending the items

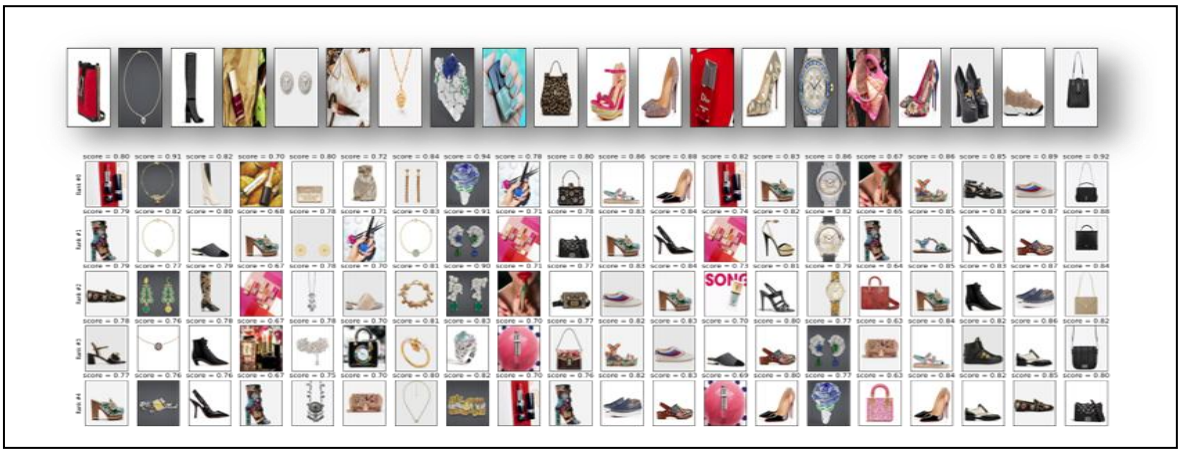

Figure 12. Multiple images as input and corresponding output using pretrained models

## 5. Conclusion

The entire work revolves around the concept of the similarity of images and then utilising this similarity for the recommendation process. The approach-I(HOG based Feature extraction and recommdation) does not peform well and needs further improvement with respect to the features extracted and subsequently recommending new images or items. On the otherhand, the pretrained models perform well, atleast in comparison with the approach-1, but still, there are chances of further improvementby refining the image dataset and trying for newer version of the pretrained models.

## References

[1] Agarwal, C., Recommender Systems, Springer, 2016.

[2] Jagadeesh,V.,Piramuthu, R., Bhardwaj, A., Di, W., and Sundaresan, N., "Large scale visual recommendations from street fashion images,". *In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, pp. 1925–1934, 2014.

[3] Krizhevsky, A.,Sutskever, I. and Hinton, G.E., "Imagenet classification with deep convolutional neural networks". *In Advances in neural information processing systems*, pp. 1097–1105, 2012.

[4] McAuley, J., Targett, C., Shi, Q., and Hengel, A.V.D., "Image-based recommendations on styles and substitutes". In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval*, ACM, pp. 43–52, 2015.

[5] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition". *arXiv preprint arXiv*:1409.1556, 2014.

[6] Wang, S., Wang, Y., Tang, J., Shu, K., Ranganath, S.& Liu. H. *"What Your Images Reveal: Exploiting Visual Contents for Point-of-Interest Recommendation". In WWW '17: Proceedings of the 26th International Conference on World Wide Web*, ACM, pp. 391–400, 2017. https://doi.org/10.1145/3038912.3052638

[7] Chen, L., Yang, F. & Yang, H.,"Image-based product recommendation system with convolutional neural networks," Stanford University, 2017.http://cs231n.stanford.edu/reports/2017/pdfs/105.pdf
[8] Tuinhof, H., Pirker, C.&Haltmeier, M., "Image-Based Fashion Product Recommendation with Deep Learning,". *In LOD 2018: Machine Learning, Optimization, and Data Science*, Springer, pp. 472-481, 2019. https://doi.org/10.1007/978-3-030- 13709-0_4
[9] Yu, W., Zhang, H., He, X., Chen, X., Xiong, L.,& Qin, Z.,"Aesthetic-based Clothing Recommendation,", *In WWW '18: Proceedings of the 2018 World Wide Web Conference*, ACM, pp. 649–658., 2018. https://doi.org/10.1145/3178876.3186146
[10] Chu, W.T.,& Tsai, Y.L., "A hybrid recommendation system considering visual information for predicting favorite restaurants," *World Wide Web*, vol. 20(6), pp. 1313–1331, 2017. https://doi.org/10.1007/s11280-017-0437-1
[11] Li, Z., Liu, F., Yang, W., Peng, S. and Zhou, J., "A survey of convolutional neural networks: analysis, applications, and prospects". *IEEE transactions on neural networks and learning systems*, 2021.
[12] *Goodfellow, I.,Bengio, Y., Courville, A., Deep Learning, MIT Press, 2016.*
[13] Simonyan, K., Zisserman, A.,"Very deep convolutional networks for large-scale image recognition." *arXiv preprint arXiv:1409.1556* (2014).
[14] He, K., Zhang, X., Ren, S. and Sun, J.,"Deep Residual Learning for Image Recognition,"*IEEE Conference on Computer Vision and Pattern Recognition*, June 27-30, 2016. Pp. 770-778, IEEE Xplore. [https://iq.opengenus.org/resnet50-architecture/]
[15] Bansal, M., Kumar, M., Sachdeva, M., *et al.*" Transfer learning for image classification using VGG19: Caltech-101 image data set," *Journalof Ambient Intelligence and Humanized Computing*, 14, 3609–3620 (2023). https://doi.org/10.1007/s12652-021-03488-z
[16] Chollet, F., "Xception: Deep Learning with Depthwise Separable Convolutions," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 1800-1807, doi: 10.1109/CVPR.2017.195.
[17] Dong, K., Zhou, C., Ruan, Y.& Li, Y.,"MobileNetV2 model for image classification,"*In 2020 2nd International Conference on Information Technology and Computer Application (ITCA)*, pp. 476-480, 2020. IEEE.
[18] Dalal, N. and Triggs, B., "Histograms of Oriented Gradients for Human Detection", *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, ***20-25 June 2005,*** San Diego, CA, USA.
[19] Style Color Image Dataset. Source: https://www.kaggle.com/datasets/olgabelitskaya/style-color-images
[20] Chen, L., Yang, F., & Yang, H. (2017). Image-based product recommendation system with convolutional neural networks. Stanford University. http://cs231n.stanford.edu/reports/2017/pdfs/105.pdf

✔