# Recommendation System using Content Based Visual Similarity

Aashutosh Marathe, Maheshwari Patil, Shoeb Shaik, Snehal Waghmare and Kalpana Metre



January 25, 2021

# Recommendation System using Content Based Visual Similarity

**Aashutosh V. Marathe[1], Maheshwari K. Patil[2], S. R. Shaik[3], Snehal Waghmare[4], K. V. Metre[5]**
**MET's Institute of Engineering, Nashik, India**
**[1]maratheaashutosh2013@gmail.com, [2]maheshwarikpatil244@gmail.com, [3]Shoebsk2410@gmail.com, [4]Waghmaresnehal897@gmail.com, [5]kvmetre@gmail.com**

***Abstract* –A smart search engine presented for online shopping. Basically it uses images as its input, and tries to understand the information about products from these images. First use a neural network to classify the input image as one of the product categories. Then use another neural network to model the similarity score between pair images, which will be used for selecting the closest product in our e-item database. Recommendation systems (RS) play an important role in e-commerce applications as they help the consumers in choosing the required items within reduced time. The traditional methods of collaborative filtering, fail to capture the visual data associated with the items. Visually-aware recommendation systems are up coming in e-commerce applications that use the visual features of the products rather than the user profiles. Deep learning techniques are used for the classification and prediction in visual recommendation systems.**

***Keywords — Image color analysis, Image retrieval, Deep learning, image Similarity, recommender Systems,***

## 1. INTRODUCTION

The design of such recommendation engines depends on the domain and the particular characteristics of the data available. Recommender systems will develop to help close the gap between information collection and analysis by filtering all of the available information to present what is most valuable to the user. The goal of a Recommender System is to generate meaningful recommendations to a collection of users for items or products that might interest them[1]. Suggestions for books on Amazon, or movies on Netflix, are real world examples of the operation of industry-strength recommender systems. The major aims of recommender systems are helping customers to find products which they seek, and the other one is increasing sales ratios for sellers. This recommendation algorithm is based on colored-feature extraction of images in order to demonstrate impressions related to the human perception of images. This system retrieves and ranks the images corresponding to the desired impressions with utilizing extracted color features[2]. There are two basic steps for CBIR process: Feature vector extraction of each image and similarity calculation between the image vectors it is one of the most challenging tasks to provide a fast, accurate and efficient model to extract an automatic feature vector of target images. Another challenge is to label a large amount of training data. To overcome these limitations, the interest towards semi-supervised and unsupervised learning techniques has increased [3]. Our main contributions are development and evaluation of deep learning based image retrieval techniques by providing comparative for shoe fashion items recommendation.

## 2. RELATED WORK

The goal of content-based recommendation system is to retrieve and rank the list of items that are closest to the query item. Today, almost every e-commerce platform has a recommendation system strategy for products that customers can decide to buy. describe our work on creating a Generative Adversarial Network based image retrieval system for e-commerce platforms to retrieve best similar images for a given product image specifically for shoes[4]. Compare state-of-the-art solutions and provide results for the proposed deep learning network on a standard data set.

Due to the great interest in e-commerce applications, recently image retrieval techniques for fashion items have gained huge popularity. Many studies have used machine learning and predictive analysis to retrieve and recommend fashion items for each customer.

## 3. LITERATURE SURVEY

In follow research paper recommender system is used for the recommendation purpose and easy availability of data. Research publications on GAN have started in 2016 and gained momentum in 2017. Alec Redford proposed convolution nets, as used in deep learning, both for the generator as well as the discriminator and applied for image representations. Centered on using CNNs which are used in supervised learning, for unsupervised learning tasks[15]. Marco Marchesi investigated DCNNGAN for generating high quality mega pixel images. Limited data is used as opposed to thousands of images used by the other researchers[16]. Hamid Eghbalzaseh introduced a general likelihood estimation for assessing the quality of generated images using GAN[17]. Scholarly search engines such as Google Scholar focal point on typical text mining and citation count Each concept does have disadvantages, which restricts its correctness for generating recommendations. In e-commerce shops that contain a enormous amount of items and users there are almost forever users that contain ratings for a few items. Using shared and other approaches proposed systems generally build neighborhoods of users using their profiles. If a user have examined just little items then it's appealing hard to conclude his taste and she/he can be associated to the incorrect neighborhood as sparsity[9]. Privacy have been the most vital problem. In order to obtain the most correct and accurate suggestion, the system must obtain the mainly quantity of information probable about the user, including data about the location of a specific user and demographic data[20]. Naturally, the question of security, reliability and confidentiality of the given data arises. Various online shops offer efficient security of privacy of the users by using dedicated algorithms and programs[22].

## 4. PROPOSED ARCHITECTURE

The ability of machines to generate pictures has been improving way beyond expectations over the course of the last decades. Such improvements have led researchers to analyses the techniques used to generate these images, thus allowing a more precise classification system that will be used here to explain the separate approaches to image synthesis through machine learning. According to "Image Synthesis Using Machine Learning Techniques".

**Feature Vector (FV) Service**: The FV Service extracts the last layer features (embeddings) from our CNN model. Despite the relatively lower performance, we opted for multi-CPU based inference to be horizontally scalable in a cost effective manner. An Elastic Load Balancer routes requests to these CPUs optimally. Additional machines are added or removed depending on the load.

**Nearest Neighbor (NN) Search**: The most computationally intensive component of the Visual Recommendation System is the k-Nearest-Neighbor Search across high-dimensional (4096-d) embeddings. Approximate nearest neighbor techniques like Locality Sensitive Hashing (LSH )

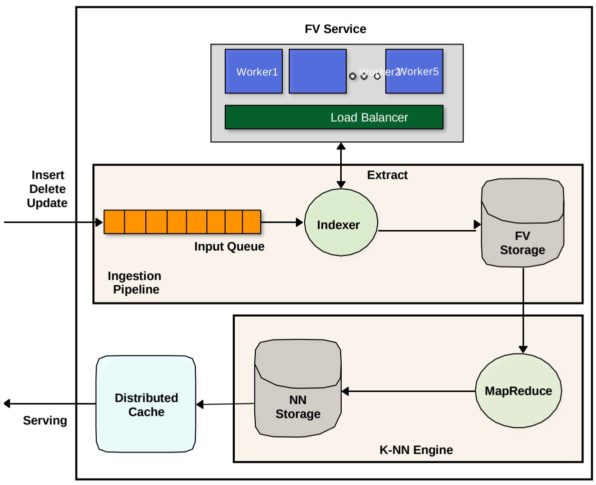


**Fig.1 : System Architecture**

significantly diminished the quality of our results (by over 10%), and K-d trees did not produce a performance gain since the vectors are high dimensional and dense. Hence, we resorted to a full nearest neighbor search on the catalog space. As can be seen in Table 4, computing the top-k nearest neighbors for 5K items across an index of size 500K takes around 12 hours on a single CPU[2]. While this can be sped up by multithreading, our scale of operation warrants the k-NN search to be distributed across multiple machines.

Our k-NN engine is a Map-Reduce system built on top of the open-source Hadoop framework. Euclidean Distance computation between pairs of items happens within each mapped and the ranked list of top-k neighbors is computed in the reducer. To make the k-NN computationally ancient, we make several optimizations.

1) **Incremental Map Reduce updates**: Nearest neighbors are computed only for items that have been newly added since the last update. In the same pass, nearest neighbors of existing items are modified wherever necessary.

2) **Reduction of final embedding size**: As can be seen from k-NN search over 512-d embeddings is significantly faster than over 4096-d embeddings. Hence, we reduced the final embedding size to 512-d by adding an additional linear-embedding layer (4096x512) to VisNet, and retraining. We observed that this reduction in embedding size dropped the quality of results only by a small margin (around 2%).

3) **Search Space pruning**: We effectively make use of the Meta- data associated with catalog items such as vertical (shirt, t-shirt), gender (male, female) to significantly reduce the search space.

While these optimizations make steady-state operation feasible, the initial bootstrapping still takes around 1-week. To speed this up, we developed a GPU-based k-NN algorithm using the CUDA

A. Dataset Processing and Data Clearing

1) Dataset Processing:
2) Dataset Clearing:

Transfer learning By performing generating image embedding using a pre-trained network such as VGG19. This is done by removing its last few layers, and performing inference on our images vectors for the generation of flattened embedding. No training is needed throughout this entire processing, only the loading of the pre-trained weights.


**Fig.:2. Dataset Processing**

Training Auto encoders Train both a simple auto encoder and a convolution auto encoder on our database images with the objective of minimizing reconstruction loss.

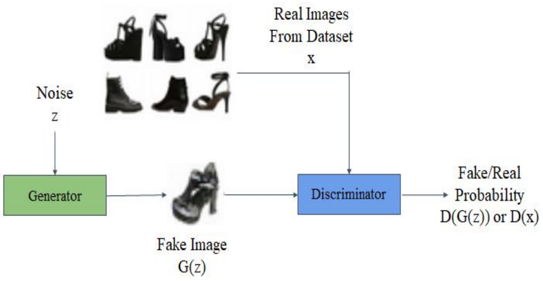


**Fig.:3 Dataset Clearing**

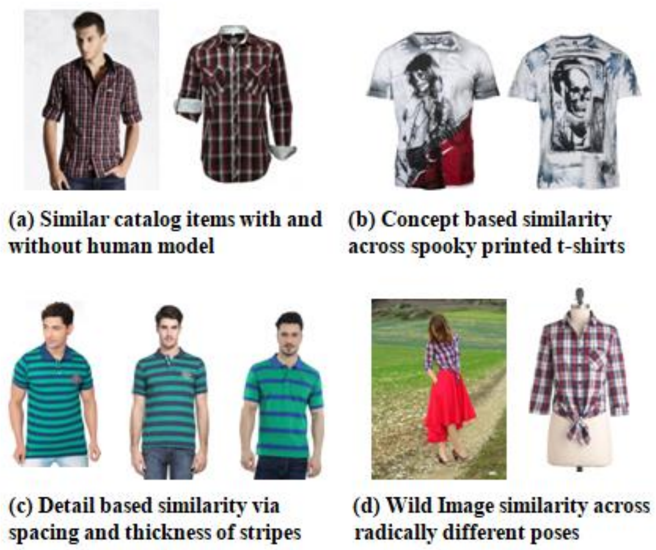


**Fig.:4. Visual Similarity Challenges**

Training Data Set is the first step for training our model, as can be easily guessed, is to acquire suitable training data out of a game. Luckily this has been kindly provided by Microsoft, using game save data from fifty Mine craft worlds generated through project Malmo, a particular research project centered on artificial intelligence. Specifically, a boot has been sent to travel around each of these worlds, filling the missing chunks as it walked. This raw data has then been processed through open-source software called Map crafter (Moritz Hilscher, 2018). Map crafter is a powerful tool that can generate isometric zoom able views of the processed maps, navigable through a browser-ready HTML file.

The recommendation system using Generative Adversarial Networks is used. Here it refers to how hypothesis about how recommendations could be effectively generated. To differentiate how specific the idea is, distinguish between recommendation classes, approaches, algorithms, and implementations.

A recommendation algorithm precisely specifies a recommendation approach. Finally, the implementation is the actual source code of an algorithm that can be compiled and applied in a recommender system. To the traditional methods of collaborative filtering, fail to capture the visual data associated with the items. The modern and faster technique of recommendation through graphical means can help the person for remembering the similar content for larger time. To design recommendation system guides the users to make decisions based on the visual features of the image. The similar set of images can be given to user and recommend him about the object whether user have any idea about the object before. It also helps the user to choose the required item faster than normal time required when user use traditional or simple technique.

### APPLICATIONS OF GAN'S

Discovering new applications for adversarial training of deep networks is an active area of research. Examine a few computer vision applications that have appeared in the literature and have been subsequently refined. These applications were chosen to highlight some different approaches to using GAN-based representations for image-manipulation, analysis or characterization, and do not fully reflect the potential breadth of application of GANs.

Using GANs for image classification places them with in the broader context of machine learning and provides a useful quantitative assessment of the features extracted in unsupervised learning. Image synthesis remains a core GAN capability, and is especially useful when the generated image can be subject to pre-existing constraints. Super-resolution offers an example of how an existing approach can be supplemented with an adversarial loss component to achieve higher quality results. Finally, image-to-image translation demonstrates how GANs offer a general purpose solution to a family of tasks which require automatically converting an input image into an output image.

These operators handle the change in sampling rates and locations, a key requirement in mapping from image space to possibly lower- dimensional latent space, and from image space to a discriminator. Further details of the DCGAN architecture and training are presented in Section IV-B. Wu et al synthesized novel objects including chairs, table and cars; in addition, they also presented a method to map from 2D image images to 3D versions of objects portrayed in those images.

### ADVANTAGES OF GAN'S

1. GANs are an unsupervised learning method: Acquiring labeled data is a manual process that takes a lot of time. GANs don't require labeled data; they can be trained using unlabeled data as they learn the internal representations of the data.

2. GANs generate data: One of the best things about GANs is that they generate data that is similar to real data. Because of this, they have many different uses in the real world. They can generate images, text, audio, and video that is indistinguishable from real data. Images generated by GANs have applications in marketing, e-commerce, games, advertisements, and many other industries.

3. GANs learn density distributions of data: GANs learn the internal representations of data. As mentioned earlier, GANs can learn messy and complicated distributions of data. This can be used for many machine learning problems.

4. The trained discriminator is a classifier: After training, system gets a discriminator and a generator. The discriminator network is a classifier and can be used to classify objects.

## 5. CONCLUSION

Extraction technique system is done on the use of datasets which will help to retrieve the data and will help to match with input by client. Here the automatically extracted visual features of image or text in the Recommender System. Recommendation systems have proposed a new content base recommender System that encompasses the technique to automatically analyze content of the object and to extract the similar content related to it. The recommendation can be done on the bases of pixels, colors, shape, size and many more. The mainly use of our project would be for the commercial site for the faster growth and the objects which can be visually represented can get the quicker knowledge to a client about the specific structure of the object. This System can be used to reduce the time for the client and get quicker access to the product which the client requires.

### REFERENCES


[1] Creswell, Antonia, et al. "Generative adversarial networks: An overview." *IEEE Signal Processing Magazine* 35.1 (2018): 53-65.

[2] Smeulders, Arnold WM, et al. "Content-based image retrieval at the end of the early years." *IEEE Transactions on Pattern Analysis & Machine Intelligence* 12 (2000): 1349-1380.

[3] Premachandran, Vittal, and Alan L. Yuille. "Unsupervised learning using generative adversarial training and clustering." (2016).

[4] Zhu, Jun-Yan, et al. "Unpaired image-to-image translation using cycle consistent adversarial networks." *Proceedings of the IEEE*.