Recognition and Classification of Products by Visual Search Using Deep Neural Networks

Karan veer Singh
MSc In Data Analytic
National College of Ireland
Dubin, Ireland
x20146248@student.ncirl.ie

Laiba Rehman
MSc In Data Analytic
National College of Ireland
Dubin, Ireland
x20144032@student.ncirl.ie

Aafaq Iqbal Khan
MSc In Data Analytic
National College of Ireland
Dubin, Ireland
x20108851@student.ncirl.ie

Priyanka –
MSc In Data Analytic
National College of Ireland
Dubin, Ireland
x20192037@student.ncirl.ie

***Abstract*—Recommender system is an important tool in the e-commerce industry to boost the sales of the business. Currently many techniques has been explored for recommending the products based on the textual representation of the data. However, textual representation is not an optimal way for identifying the similar products. Therefore, in this work we have explored visual based product recommendation system. Where we are utilizing multiple pre-trained deep learning models such as VGG-16, VHH-19. ResNet and Inception-V3. After certain set of experiments, we have identified that ResNet model outperforms than all the models in terms of Accuracy, loss and PRF Score.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

One of the major advantages of the digital age is the rise of the e-commerce industry. With the help of online stores and e-commerce platforms, the user can get whatever they want from the comfort of their homes. This exclusive feature has made the field boom in many ways. The greater demand has also rapidly increased the supply of the e-commerce industry. Although e-commerce is a great option for purchase and procurement, there are a very problems and challenges involved in it too. Mostly on a large scale, there are issues seen on the two main domains, which are from the perspective of the buyer and the perspective of the seller. The process of e-commerce listing and selling works in the simplest ways, and the seller clicks and uploads the images of the product on their online store platform using tags and respective labels and enables the product to show up on the search results for the customer to buy in a very hassle friendly way. Some technical and manual errors can happen in this process as these are all done by individual people. The happening of such an error will hinder the product listing on the search and may not show up for the customer. This, in turn, will lead to a decline in the range of sales of that particular product. Keeping up the demand in the right way is also very important for the e-commerce business to thrive and grow in the right way.

A. Motivation

There are some existing approaches in the field of product identification and recognition. Usually, such a process is done by analysing and figuring out the keywords or sample tags used by the customer in searching for the product. The algorithm that is used for the search is designed in a way to check out the database for the related keyword used by the customer to fetch the designed products on the listing menu. Once the products are listed on the page, then the customer can look for the right items and models that will suit their requirements in the perfect manner. The text-based search works in an average manner where the pre-requisites mentioned by the customer are the main focal points in the analysis. Only with the right keyword, it functions well, and else more products have the risk of getting left out and misplaced. It might not work effectively and include all the available products in the listing.

There are more unique and beneficial products in the market that are unaware of the knowledge of customers because of the traditional keyword search for product findings. These practices limit the ability of product visibility and branding in most e-commerce business websites. The gap in the process of effective product identification and recognition can be quickly solved by using the advanced techniques of machine learning and visual search. Visual search is nothing but an all-inclusive process of finding out the desired range of products or similar using the method of photo analysis. When a particular customer goes for a photo or visual search, unlike the keyword recognition algorithm, this model searches for the likewise and identical photos of other products that are listed on the e-commerce website by looking into the database. By just clicking and uploading the like model picture of the product they need to buy to the visual search option, the customer can also easily get immediate items that are similar to the ones they want to purchase.

B. Objectives

More high-end companies and artificial intelligence giants have their own visual search options like google lens, Visenze, etc. The product surfing using these models is more effortless and hassle free. This kind of visual search carried out by the search engines tends to produce accurate and effective results for the product search. To amplify and enhance the process is where machine learning algorithms come into the picture. It helps in prompt classification of images with excellent accuracy and makes it visible for the customer related

to their web search. The domain of the visual search and image identification and classification comes under the broad unified category of unsupervised problems. Here the machine learning algorithms can be employed to facilitate easy product recognition, classification, and recommendation. The image uploaded by the user will be considered as the target image, and the model responds by identifying and displaying similar images on the website related to the original one uploaded by the customer. The process is said to be straightforward and effective. In this research paper, we have planned to use pre-trained deep neural networks for the technique of image classification and recognition. A huge set of image data of numerous products are obtained for the classification process. The techniques of data modelling and pre-processing are carried out on the procured image sets. VGG-16, VGG-19, and RestNet Architecture are some of the pre-trained deep neural network architecture models used in this research work. In order to have an effective outcome and optimal working model of image classification, we consider the following metrics like accuracy, precision, recall, loss, and F1 score to estimate the range of performance of the designed model.

C. Research Question

- Which Deep Neural network algorithm optimally recommends the similar product images ?
- Does text-based recommendation methods are always effective for product recommendation ?

II. Literature Review

This section of the research work deals with the complete insight and analysis of the existing model used for the image recognition, classification, and identification. A detailed examination of various visual search models is also presented in this literature review.

A. Using Pre-Trained Deep Neural Networks

A very extensive research work performed on the visual search technique employs the use of pre-trained deep neural networks. In the research model, the fashion image classification is taken for the problem analysis. The author presented the various challenges and image identification issues experienced by the concerned e-commerce industry. The issues of manual tagging and product listing are also taken into consideration by the author here. The risk of potential bottlenecks and the rise and fall of the demand and sales in the product listing is also mentioned in the paper by the author [1].

The authors also lay needed emphasis on the working process and effectiveness of machine learning algorithms employed in the approach. Same as this paper, the mentioned research also uses the architecture of VGG-16 for image classification. The results of this research work are seen to be much better than the keyword analysis model in terms of product recognition and listing. Many a times there are problems that arise in the process of image classification. It can happen with the visual search too. In the work presented here, the author tries to solve the issues that occurs in the process of visual and image recognition. For this work, the publicly available dataset called CheXpert is taken into consideration by the author. It contains the whole list of automated chest x-rays. This analysis carried out here will showcase the disease condition using the automated image classification [2]. Here in this research the process of under sampling and weighted class approach is widely carried out. This research work is based on the Synthetic Minority Over sampling technique called as SMOTE as the underlying functional mechanism. The data augmentation is the main and predominant problem that arise in the classification technique. The employment of such algorithms helps in lowering such issues from happening. The complete technique of data augmentation is given in the work done by Brownlee. The author complies and gathers all the information and past research work present in this domain to perform the following analysis. In this research the author also explains all the in-depth problem and issue with the augmentation of the images and how it affects the process of visual search to bring the product out in the listing. All kind of validated, pre-tested and raw dataset samples are used in this work [3]. All kind of datasets are used here to check the effectiveness of the model. This complete configuration guide helps the users or online business store owners to go beyond the technical issues in having the right visual search protocol in place.

Fig. 1. Architecture of VGG16 model

In another research that involves the method of image classification and recognition, the author Ding uses various methods in place to obtain the desired results and outcome. The employment of deep learning algorithms like deep neural network are also done in this research and analysis work. In this work, the author presents the model that is capable of identifying the presence and the various stages of Alzheimer's disease [4]. Convolutional neural network is also presented in this research work. The outcomes of the model portray the accuracy around 85 percentage and the image listing is also done in a perfect manner. Since huge number of images of involved in this approach, a clarity is needed to handle the method and process in the right manner. Such a work is presented by Seif, in this work. The author states some of the techniques and effective ways to handle the imbalanced datasets. The author uses the technique of class balancing and performing methods to tackle such complicated and

technical problems [5]. The emphasis of using the minority class is also mentioned in the research work. The author also explains about all the criteria like weight balancing, over and under sampling and other highlights. The results of the work will help the user or e-commerce expert in understanding the visual search method in a better manner. Another approach that involves the use of machine learning algorithms for the visual search is carried out by Umar. In this paper, the authors perform this method for the classification and recognition of cosmetic products. Feature extraction is a very important process carried out in this research [6]. For the testing and processing, the author makes use of the algorithms like support vector machine, KNN, decision tree, and artificial neural networks. Some of the very important tasks like customer base analysis and brand recognition process is also carried out in this paper [7]. The data used for this research is obtained from the trusted dataset called as Kaggle. The outcomes of this research are found to be comparatively better than the word-based analysis. The cosmetic product based visual search has the accuracy of around 90 percentage. One of the very important process of e-commerce section is the numerous product recognition and analysis process. Since there are tons of products, the method of employing visual based recognition works well than the keyword based one. Even the tiny and the little differences are observed and the classification and identifications are done in an accurate manner.

Multi-learning algorithms are used for the employment of the model [8]. All the product images are listed and mentioned in the clear way with all the required labels and indexes. All the little steps of the designed model are explained clearly in the paper and the eProduct listing are listed in a very clear way. The queries and the indices are also mentioned in the right way. All the stipulated results are mentioned in the clear manner. Another such research done by a scholar mentioned the use of deep learning algorithms and convolutional neural networks for the employment of fashion image classification. Like most models, here also the authors make the model in the way the image listing on the e-commerce website must be in a way according to the image presented in the visual search option by the customer. Here the hierarchical structure is considered for the enhanced working of the model. Using the architecture of VGGNet the use of CNN is proposed in the paper [9]. The outcomes of the mentioned model show the better function of the image classification and recognition. The accuracy and enhanced detection are better in the visual search here than the other practised models. Another such research work that deals with the object classification for the domain of fashion products is presented in this paper. Here the author employs the techniques of convolutional neural network for the classification and recognition process. The division and separation of the outfits are divided according to the type and model of the clothing that are mentioned in this paper in a very detailed manner [10]. A very extensive dataset with all the sample images is also included in the method. The outcomes of this research prove the use of convolutional neural network in the image analysis found to work better than the most models present in the market. The use of convolutional neural network is seen to be very advantageous and beneficial for the e-commerce business owners in the process of visual recognition and analysis. One such theory is proved in the work carried out by research done on the image classification and image retrieval process.

The use of AlexNet is also seen to be employed in the mentioned process. In the research presented here, four classified are used to enhance and make the process better [11]. Trained on a very advanced model using the graphic cards and automated process, the employed model will reduce the happenings of human errors that occurs due to involvement of people. Extensive experiments are carried out here to prove the well working criteria of the model. The outcomes of this research prove to be effective than the existing models. A very different approach of visual image recognition is carried in the research performed by Cheng. The author employed the deep learning technique for the image classification, and recognition of the abdominal ultrasound images. The convolutional neural networks are applied in this research work to study all the sets of categorial images taken for analysis [12]. The layers of architecture included here is the VGGNet and CaffeNet. Feature extraction is another important step carried out in the image recognition work. The accuracy and performance of both the models are found to be well working and has improved results. The accuracy range is above 90 percentage.

B. Using MobileNet architecture

The use of visual image recognition has become one of the most advantageous processes for all e-commerce stores and business owner. The process of product search and image listing has become so straightforward and easy. Here the author uses the multi-layer convolutional neural network for the analysis. For the image classification and identification, the MNIST dataset is used in this approach [13].

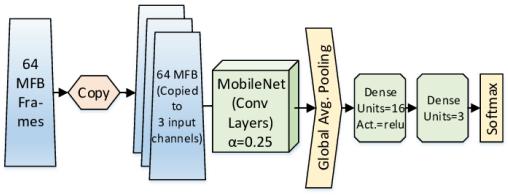

Fig. 2. Architecture of MobileNet

The feature extraction is employed as a very important step for the enhanced performance and accuracy. One of the essential things mentioned by the author in the approach of performing the good image recognition process is to choose

the right deep learning algorithm for the model. The outcomes prove that the use of deep learning methodology gives out impressive results.

Another similar research work that deals with the fashion image classification and recognition is found to be presented by Xuan. Here the author employs the different algorithmic models of convolutional neural network and transfer learning [14]. This is done in the intention of improving the performance standards and the range of accuracy. The categories of the image samples are divided into eight different types and then the examination is properly initiated. For the analysis here, the street style fashion dataset is taken as the input sample dataset unit. Some of the pre-trained models are also taken into consideration here. The results of the experiment show considerable improvement in the classification and recognition of the image appeared in the visual search. From all the above-mentioned existing research work, it is established that the use of deep learning algorithms and neural networks are very effective and well working in the analysis of identifying and recognition images in the visual search.

To prove that statement one such research is performed by Di [15]. Here the authors make a comparative study of all the deep learning algorithms and neural networks employed in the process of image recognition and visual search. The same fashion MNIST dataset is taken for the examination here also. The author also employs various algorithms and principles in this research work. The outcomes establish the use of MobileNet model is more effective than all other traditional deep learning models used in this approach [16]. The structure is found to be very simpler and it also renders higher standards of accuracy and greater performance. To see the effectiveness of some other typical approaches for the same product listing and image classification a new research sheds more light on the topic. The hierarchical model combined along with the algorithms of deep learning convolutional networks provides more answers for the same. From lower layer to the top layer every single thing is taken into consideration in this research [17].

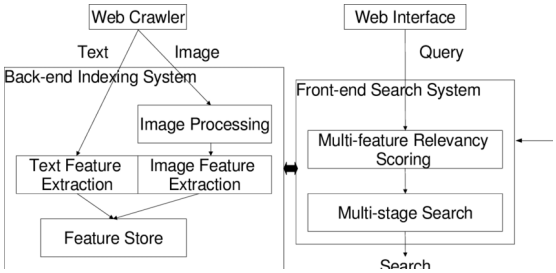

Fig. 3. Architecture of the visual product image search system

This represented model is nothing but an extensive knowledge-based classifier approach for the hassle-free image recognition process. The VGGNet framework is also found to be used here in this approach. Additionally, the well-known MNIST dataset is also taken for analysis in this approach. The results seen in this approach is found to be very improved and beneficial. As fashion industry is the most booming sector of the e-commerce domain, the performance and product listing must be every excellent and on spot [18]. It is also seen that the visual search tends to pitch in more sales by making the product appear on the page when the customer search for something related. To enhance the computer based vision search technique, the presented paper uses the prevalent technique of auto encoding, separation, classification and listing of images using the algorithms of Deep learning ConvNet technique [19]. As usual like other research works, the method of feature extraction is found to be the major underlying step in this analysis. The image is likewise embedded properly and the pixel wise grouping and allocation is also done in a wise manner. The segmentation is carried out by the mask RCNN, which makes the process simpler and even more effective. In this research work, the main thirteen classes and features are considered to analyse the performance [20]. The outcomes establish the better accuracy, improved search results and reduced loss.

III. Methodology

The main objective of this research is to recommend the similar products to the customers using different methods. In this work, we will utilize two different types of methods for product recommendation, the text-based recommendation and image-based recommendation. Our proposed methodology consists of multiple steps, the includes the data pre-processing, text pre-processing, Model training, evaluation etc. We will review in-depth analysis about each step in further subsections.

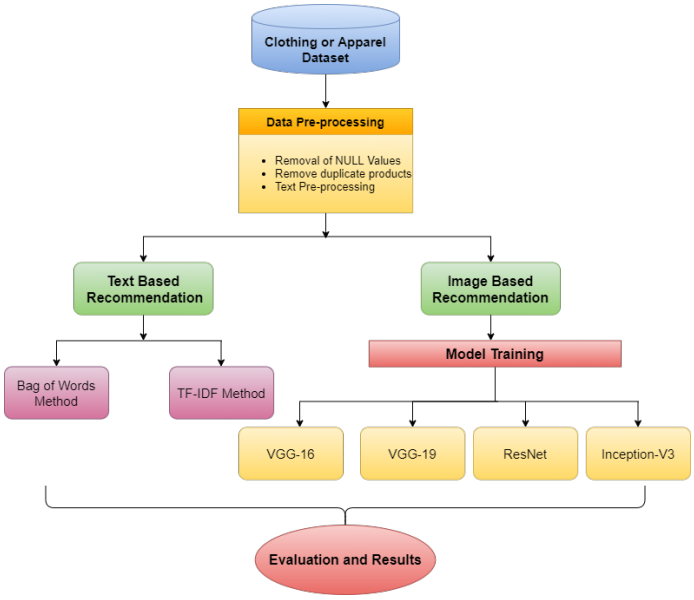

Fig. 4. Proposed Method for Product Recommendation

A. Dataset Collection

The dataset has been collected from Amazon Women Apparel dataset, which has been collected from kaggle. The dataset consists of approximate 1,80,000 number of samples with 7 features. The features of the dataset contains the information such as asin, brand, color, images URL, title of the product, product type and their price. The samples of the dataset is in raw format, which needs to pre-processed for utilization in efficient way. The dataset mainly consists the information about different fashion products.

B. Data Pre-processing

In order to achieve the optimal prediction, data pre-processing is the key step. Data pre-processing in itself consists of many steps that includes handling duplicate values, removal of null values and various other text pre-processing parts. As we have described that dataset consists of 1,80,000 samples before eliminating the null values, the values obtained after eliminating the null prices and null color are 28,385. After removing the null values from the dataset, the product title having the short description will be also be removed from the dataset. Another issues with the dataset we have found that it contains the duplicate title. In order to achieve the better results, we have removed the duplicate title from the dataset. The total number of duplicate titles are found to be 2325, which has been removed from the dataset. After removing the duplicate titles the total number of data samples are found to be around 17,593. After this text processing technique has been applied, using natural language processing techniques. The implementation of various operation of natural language processing has been applied using nltk library. For the text-preprocessing we have applied the operations such as removal of stop words, punctuations or special characters etc. On the above described dataset, the text pre-processing step was executed in 6 seconds. The next task is to extract the images from the URL provided in the dataset, which will help us to identify the recommendations.

C. Recommendations

In this research, the products will be recommended based on the image and text based recommendation method. The explanation of the each method has been discussed in further subsections.

1) Text-Based Recommendation: For the text-based recommendation we are going to use two different method. The methods are Bag of Word (BoW) method and Term Frequency — Inverse Document Frequency (TF-IDF) method. The performance of these methods can be compared by calculating the Cosine distance for identifying the similarity. Based on the Cosine distance, the pairwise Euclidean distance of method can be calculated or has been sorted in the increasing order. The method calculates the minimum euclidean distance will be considered as an optimal method for text based recommendation. We will discuss about the recommendation results of these methods in results and evaluation section.

2) Image-Based Recommendation: For the Images based recommendation, all the images needs to be extracted from the URLs. Once the images are extracted the pre-processing needs to be performed. On analysing the label of the dataset, it has been found that the dataset is highly imbalanced in nature. There are certain categories of the products which contains the minimum number of samples, other than the top 3 categories all the remaining categories are put up into the new category named as Other Category. In order to generate the biased results for our analysis, we have balanced the dataset and categorized it in mainly the 4 categories. Category name with number of samples is shown in Figure 5.

APPAREL	1500
BOOKS_1973_AND_LATER	1500
SHIRT	1500
OTHER	1500

Fig. 5. Balanced dataset based on Product Type

After successfully balancing the dataset, pre-processing needs to be applied over the images. In the first step, each image is reshaped to the size of 128 X 128. In the dataset, some black & white images are found which needs to be converted into RGB format. The labels of the each image is encoded using One-hot encoding. Using One-hot encoding, each label will be converted into sparse array. In order to perform the predictive analysis based on the certain product images, the dataset is splitted into training and test set. In this work, we have utilized the 70% of data for training whereas the 30% of the image samples will be used for testing. We have utilized the multiple Pre-trained models based on deep neural network architecture. The models that will be utilized in this research are VGG-16, VGG-19, ResNet and inception-V3. In order to reduce the complexity of input features, we have added a some layers of dense network where we are also utilizing the batch normalization and dropout function. The model is trained over the 5 number of epochs where the batch size is selected as 4. The proposed deep neural network architecture consists of input layer, 2 hidden layers and output layer. For hidden layers, we are using the ReLu as the activation function and for output layer, softmax is used as an activation function.

D. Model Evaluation

As we are using text based and image-based recommendation methods. The evaluation metrics for each method will be different from one another. For each method an evaluation will be performed among multiple algorithms. For the text-based method we will compare the Euclidean distance among the two pairs of images, using Bag of words and TF-IDF method. The method which have the minimum Euclidean distance between

the different pairs of images, will be considered as the optimal algorithm for text-based recommendation. For both the text based algorithm we will also calculated the average euclidean distance to get the better insight about the recommendation. On the other hand, for image based recommendation system we are utilizing the 4 different deep learning models that includes VGG-16, VGG-19, ResNet and Inception-V3. Each model will be compared with each other in terms of accuracy score, loss score, Precision, recall and F1-Score. The model which will have the minimum loss and maximum accuracy with high PRF score will be considered as the optimal and best model for product recommendation based on the images.

IV. Results & Discussion

In this section, we will discuss about the results of the multiple experiments performed in this work. The experiments can be categorized into the two subsections text-based recommendation and image-based recommendation. In each method, multiple algorithms have been utilized and for each algorithm we will calculate the multiple metrics. For text-based recommendation system, the euclidean distance is measured as the performance metrics. Whereas, for image-based recommendation system we will be using pre-trained deep learning models such as VGG-16, VGG-19, ResNet and Inception-V3. For training the deep learning models the dataset has been splitted into the training and testing set with the ratio of 70:30. The outcomes of the different algorithm for recommendation will be analyzed and discussed in more detail in this chapter.

A. Text-Based Recommendation

In the first method of recommendation based on the text data we are using Bag of words methods , where the recommendation will be based on the title of the product. The Bag of words algorithm is mainly used to find out the occurrence of the word within the document. bag of words mainly consists the dictionary of the known words. Using text vectorizer, the text will converted into the vector and the similarity between the two vectors will be identified by calculating the cosine similarity. The top 10 recommendations will be represented by calculating the sorting the indices of smaller distance. The sample output generated by Bag of words method for identification of similarity is shown in Figure 6. The average euclidean distance obtained using using Bag of word method is 2.119.

Fig. 6. Bag of Words based sample output

In the second method, the TF-IDF algorithm has been applied on the product titles. where we are counting the term frequency from the text corpus. For the calculating the term frequency, we divide the number of times the word appeared in the document with total number of words in the document. On the other hand, using IDF (inverse document frequency) we identify the weight of words which are rarely occured into the text corpus. The minimum word occurrence have high IDF score. Top 10 recommendation generated using tf-idf method is shown in Figure 7. The average euclidean distance obtained using TF-IDF method is 1.055. The euclidean distance obtained using tf-idf is found to be minimum as compared to the Bag of words method. Thus, we can say that TF-IDF algorithm outperforms than Bag of words method, for product recommendation based on the textual data.

```
First one is input image and the rest 10 are recommended clothing
```

```
Title: im huckleberry doc holliday 34sleeve raglan long sleeve
Eucliden distance from the given image : 0.0
```

```
Title: im weird im unicorn 34sleeve raglan long sleeve
Eucliden distance from the given image : 1.0649487461630647
```

```
Title: moo im goat 34sleeve raglan long sleeve
Eucliden distance from the given image : 1.1335754403057734
```

Fig. 7. TF-IDF based sample output

The comparative analysis, representing the Average euclidean distance for TF-IDF and Bag of words method is also shown with horizontal bar graph in Figure 8. The average euclidean distance for top 10 product recommendation is 1.055 and 2.119 which has been calculated by TF-IDF and Bag of words method. TF-IDF average euclidean distance is minimum, hence it is better algorithm for recommendation.

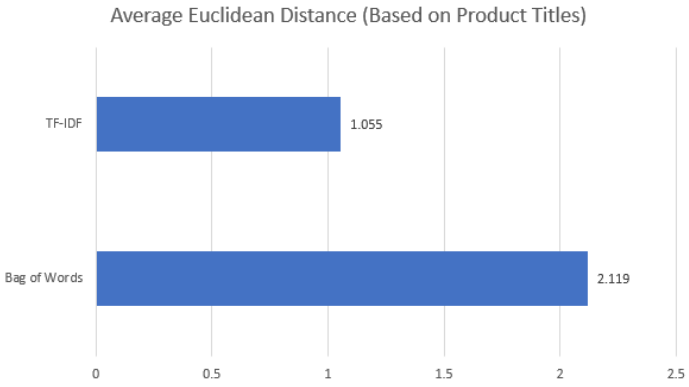

Fig. 8. Average Euclidean Distance calculation

B. Image-Based Recommendation

Previous experiment has been performed over the text-data, where the recommendation was mainly based on the title of the product. In this methods, the recommendation will be generated by analysing the certain features of the image and similar images will be predicted for recommendation. After pre-processing the image, deep learning model needs to be trained. The model generates the better accuracy, can be considered as the optimal model for product recommendation based on the image data. Other than the accuracy, we have also calculated the other score such as loss, precision, recall and F1-Score. Based on each metrics a comparative analysis between each algorithm will take place. Each model has been trained over the 5 number of epochs, where the batch size is selected as 4. The described metrics will be calculated over the test set.

1) Accuracy Comparison: Accuracy is the most important parameter to evaluate the model performance. It informs about the overall performance of the model, model having the higher accuracy score is considered as an best model for predictive analysis. In this experiment, we have calculated the accuracy score for each algorithm, the line graph representation of obtained accuracy is shown in Figure 9.

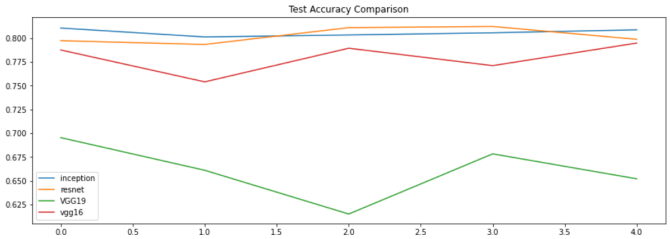

Fig. 9. Accuracy obtained over the test set

After analyzing the accuracy graph as shown in Figure 9, it has been observed that inception V3 and ResNet are the two model providing the better accuracy over the test data. The other model such as VGG-19 and VGG-16 achieves the 65.69% and 78.89%. The worst performing model is VGG-19, on the other hand, due to increasing behaviour over every epoch resnet can be considered as the best model for achieving the highest accuracy score of 82.09%.

2) Loss Comparison: Loss score and accuracy score are found to be anti-proportional to each other. If the accuracy of the model increase over every epoch then loss decrease and vice versa. The loss score of ResNet and inception-V3 model is found to be minimum and very similar. On the other hand, After 3 epoch the VGG-19 model loss score is found to be increasing in a linear fashion which representes the case of overfitting. The lines of inception V3 and Resnet are overlapping each other and having the minimum loss score.

Fig. 10. Loss score obtained over the test set

The minimum loss score obtained using ResNet and inception-V3 is found to be 1.04. The highest loss score obtained is 52.24, which is very high as compared to the other models.

3) Precision Comparison: Precision score is very important measure to evaluate the model performance, it mainly informs about the false positive values in the predictive analysis. Highest precision score represents the low false positive rates. The line graph of precision score is shown in Figure 11.

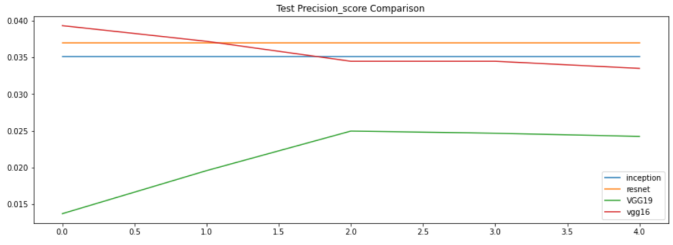

Fig. 11. Precision Score over the test set

From the graph shown in Figure 11, it is found that ResNet model represents the linear behaviour and have the highest precision score of 0.037, followed by inception-V3, VGG-16 and VGG-19. From the above analysis it is very clear that VGG-19 generates the highest number of false positive values.

4) Recall Comparison: Recall score is mainly used to indicate the false negative values in the predictive analysis. The highest recall score means the better performing model. The recall score of all the deep learning models has been represented in Figure 12.

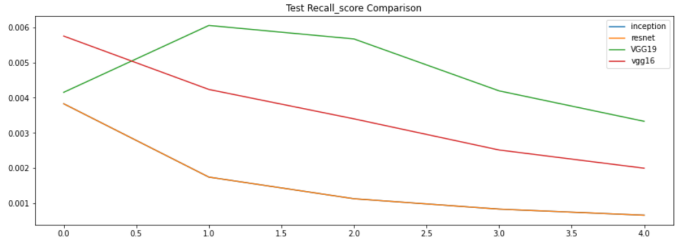

Fig. 12. Recall Score over the test set

On observing the values over every epoch it has been noted that recall score is found to be decreasing for every model. Thus, we can say that no model is performing good and model is unfitting. The reason behind the decreasing recall score is the number of training and test samples. As the total number of images samples are only 6000, which again has been divided into training and test set. Due to low number of training samples model is not able adapt the patterns from the input features. Adding more samples can solve this issue.

5) F1-Score Comparison: F1-Score is the overall score obtained from the precision and recall score. F1-Score also can be said as the combined score of precision and recall. The f1-Score obtained over the test set is shown in Figure 13.

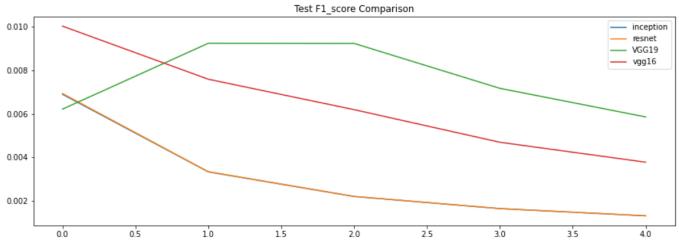

Fig. 13. F1-Score obtained over the test set

Just like the recall score, the f1-score is also found to be decreasing over every epoch. In order to improve the results, more number of image samples can be added with improved labelling.

6) Predicting the output: Based on the input image we will be predicting the 10 similar images using the best model. On analysing the accuracy, loss and precision score we have reached to a conclusion that ResNet model outperforms as compared to all the other models. A very close Accuracy, loss and precision score has been provided by the inception-V3 model which also can be considered for predictive analysis. On providing the input image, the following 10 recommendations has been generated by the ResNet Model as shown in Figure 14 and Figure 15. Further more input images can be provided in order to get the similar image recommendations.

Fig. 14. Sample Input Image

Fig. 15. Recommendation using ResNet Architecture

V. Conclusion and Future Work

In the e-commerce industry, recommending the similar products to the customers is a challenging task. The recommendation can be performed in many different way and depends on the multiple factors. In this research work, we have explored the text based and image based recommendation system. For the text-based method, using TF-IDF algorithm we were able to provide the better recommendations as compared to the Bag of words method, the euclidean distance obtained using the TF-IDF methods is found to be very minimum among the text pairs. On the other hand, for recommending the products based on the images we have utilized the 4 differernt deep learning models. After analysing the results we can reach to a conclusion that, ResNet model outperforms than all other models in terms of accuracy, loss and Precision Score and VGG-19 does not perform well for ecommerce dataset. Currently, the research has been performed on very small number of image sets. Large image set of fashion products with improved labelling can improve the performance of the model to a greater extent. In future work, the big data in cloud computing environment can be explored for handling the large volume of image and text data using Hadoop and Spark Framework. Currently, almost all the processing has been performed over single machine, utilizing the concept of hadoop and spark can reduce the computation time and model training time to an great extent.

References

[1] F. Li, S. Kant, S. Araki, S. Bangera, and S. S. Shukla, "Neural Networks for Fashion Image Classification and Visual Search," *arXiv:2005.08170 [cs]*, May 2020. arXiv: 2005.08170.

[2] A. Bhattacharya, "How to use SMOTE for dealing with imbalanced image dataset for solving classification problems," Apr. 2021.

[3] J. Brownlee, "How to Configure Image Data Augmentation in Keras," Apr. 2019.

[4] Y. Ding, J. H. Sohn, M. G. Kawczynski, H. Trivedi, R. Harnish, N. W. Jenkins, D. Lituiev, T. P. Copeland, M. S. Aboian, C. Mari Aparici, S. C. Behr, R. R. Flavell, S.-Y. Huang, K. A. Zalocusky, L. Nardo, Y. Seo, R. A. Hawkins, M. Hernandez Pampaloni, D. Hadley, and B. L. Franc, "A Deep Learning Model to Predict a Diagnosis of Alzheimer Disease by Using 18F-FDG PET of the Brain," *Radiology*, vol. 290, pp. 456–464, Feb. 2019.

[5] G. Seif, "Handling Imbalanced Datasets in Deep Learning," July 2021.

[6] Z. Zhao, H. Zhang, H. Sun, and B. Qiao, "Commodity Image Retrieval Based on Convolutional Neural Network and Late Fusion," pp. 90–100, May 2021.

[7] S. Umer, P. P. Mohanta, R. K. Rout, and H. M. Pandey, "Machine learning method for cosmetic product recognition: a visual searching approach," *Multimedia Tools and Applications*, June 2020.

[8] J. Yuan, A.-T. Chiang, W. Tang, and A. Haro, "eProduct: A Million-Scale Visual Search Benchmark to Address Product Recognition Challenges," *arXiv:2107.05856 [cs]*, July 2021. arXiv: 2107.05856.

[9] Y. Seo and K.-s. Shin, "Hierarchical convolutional neural networks for fashion image classification," *Expert Systems with Applications*, vol. 116, pp. 328–339, Feb. 2019.

[10] B. Lao and K. Jagadeesh, "Convolutional Neural Networks for Fashion Classification and Object Detection," p. 7.

[11] D. Giveki, A. Shakarami, H. Tarrah, and M. Soltanshahi, "A new method for image classification and image retrieval using convolutional neural networks," *Concurrency and Computation Practice and Experience*, July 2021.

[12] P. M. Cheng and H. S. Malhi, "Transfer Learning with Convolutional Neural Networks for Classification of Abdominal Ultrasound Images," *Journal of Digital Imaging*, vol. 30, pp. 234–243, Apr. 2017.

[13] S. Shubathra, P. Kalaivaani, and S. Santhoshkumar, "Clothing Image Recognition Based on Multiple Features Using Deep Neural Networks," in *2020 International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 166–172, July 2020.

[14] X. Xuan, R. Han, S. Ji, and B. Ding, "Research on Clothing Image Classification Models Based on CNN and Transfer Learning," in *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, vol. 5, pp. 1461–1466, Mar. 2021. ISSN: 2689-6621.

[15] W. Di, "A comparative research on clothing images classification based on neural network models," in *2020 IEEE 2nd International Conference on Civil Aviation Safety and Information Technology (ICCASIT*, pp. 495–499, Oct. 2020.

[16] A. Dagan, I. Guy, and S. Novgorodov, "An Image is Worth a Thousand Terms? Analysis of Visual E-Commerce Search," pp. 102–112, July 2021.

[17] Y. Liu, G. Luo, and F. Dong, "Convolutional Network Model using Hierarchical Prediction and its Application in Clothing Image Classification," in *2019 3rd International Conference on Data Science and Business Analytics (ICDSBA)*, pp. 157–160, Oct. 2019.

[18] X. Lin, B. Gokturk, B. Sumengen, and D. Vu, "Visual search engine for product images," p. 22, Jan. 2008.

[19] H. Jhaveri, D. Sheth, and K. Bhowmick, "Autoencoder Embedding, Segmentation and Classification of Apparel Images using a Deep ConvNet Approach," in *2020 4th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, pp. 974–979, Nov. 2020.

[20] J. Li, H. Liu, C. Gui, J. Chen, Z. Ni, N. Wang, and Y. Chen, "The Design and Implementation of a Real Time Visual Search System on JD E-commerce Platform," pp. 9–16, Dec. 2018.