International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST

Fashion Outfit Recommendation Based on Deep Learning Model

Bhagyshree Pravin Bhure[1], Pratiksha Tulshiram Bansod[1], Monali Shivram Amgaokar[1], Savita Pralhad Lodiwale[1], Anjali Pravin Orkey[1], Ashish Mohod[2]

[1]BE Student, Department of Computer Science and Engineering, Priyadrashini JL College of Engineering, Nagpur, Maharashtra, India
[2]Assistant Professor, Department of Computer Science and Engineering, Priyadrashini JL College of Engineering, Nagpur, Maharashtra, India

Article Info
Volume 8, Issue 3
Page Number : 216-225

Publication Issue
May-June-2021

Article History
Accepted : 10 May 2021
Published : 20 May 2021

ABSTRACT

Nowadays, the fashion industry is moving towards fast fashion, offering a large selection of garment products in a quicker and cheaper manner. Machine learning is completely changing the trends in the fashion industry. From big to small every brand is using machine learning techniques in order to improve their revenue, increase customers and stay ahead of the trend. People are into fashion and they want to know what looks best and how they can improve their style and elevate their personality. Traditional recommendations for clothes consisted of lexical methods. However, visual-based recommendations have gained popularity over the past few years. This involves processing a multitude of images using different image processing techniques. In order to handle such a vast quantity of images, deep neural networks have been used extensively. With the help of fast pre trained algorithms, these networks provide results which are extremely accurate, within a small amount of time. However, there are still ways in which recommendations for clothes can be improved. In this paper, we propose a deep learning based system which operates as a personal assistant to a fashion user. The system's architecture and all its components are presented, with emphasis on the data collection and data clustering subsystems. In our use case scenario, datasets of garment products are retrieved from Kaggle website which contains different outfit which includes t-shirt, shirts, shoes, dress, etc nearly 20 + categories. With this dataset, we trained our CNN model and the performance of it is over 84%. And our model can also recommend daily outfit to users.

Keywords : Outfit recommendation, convolution neural network, embedding, clustering.

I. INTRODUCTION

Fashion has a tremendous impact on our society [1]. Clothing is the most important thing that people should consider every day. People want to look good, as we can see that there are lots of online social sites such as Instagram or Facebook where people can show their fashion pictures to the world. Also

Deciding what clothes to wear for an event can often be a time-consuming task. Also people like to buy lots of clothes to try different styles to make them look more appealed, which is reflected in the growing online clothing sales, reaching 370 billion dollars in the US by 2017, and 191 billion euros in Europe [2], also over 103 billion RMBs just on one festival day in Taobao.com, an E-commerce website in China. Since fashion is such a not only popular but also extremely complex topic in our daily life, it is important for us to study about it. However, we cannot make it manually. Recent years more and more researchers started to use computer vision to study on fashion [3, 4, 5, 6]. One of the main task is application domain of fashion recommendation [7,8]. The main focus is to parse clothing from photographs, which is widely used on online shopping sites. In addition, it will make people's life more convenient if there is a system which can suggest good outfit to users based on the clothing that the users already have. An occasion-oriented clothing recommendation is given to suggest the most suitable clothing from the user's own clothing photo album [9], which just put attention to the occasion not the outfit itself.

In fashion sales, the recommendation technology, as an emerging technology, has attracted wide attention of scholars. As is widely known, the traditional garment recommendation depends on manual operation. To be specific, salesmen need to recommend garment to customers in order to arouse their interest in purchasing. However, it is very difficult for salesmen to understand customers' real thoughts and then recommend the targeted garment as there is no sufficient cohesiveness between customer information and merchants. Therefore, it is essential and meaningful to find a set of objective indicators, instead of subjective opinions, to evaluate the fashion level in the clothing recommendation technology.

Recommender systems generate greater revenue for ecommerce websites if the recommendations are good. Hence, while building such a system for garments, recommendations need to be good. Object detection is an important part of visual fashion recommendation. Traditional methods for object detection included obtaining feature descriptors like Histogram of Oriented Gradients (HOG), Speeded Up Robust Feature (SURF), Scale Invariant Feature Transform (SIFT), etc. More recently, deep neural networks have been used extensively for object detection. Artificial neural networks are computing systems which are derived from the functioning of the human brain, particularly the nervous system [2]. The main processing units are neurons, which are connected by links known as synapses. These neurons are divided into various layers. Input data is fed to these neurons in different layers which perform computation on it. There is one input layer and one output layer and multiple hidden layers. If there exists more than one such hidden layer, then the artificial neural network is called a deep neural network. Deep neural networks are much harder to train than normal neural networks. Deep neural networks are used extensively in image processing and could be used to detect objects in images. Using these, we can learn which outfits are worn at which event or used in which scenario. This will help in making recommendations which are tailored to required occasions.

In this paper, our goal is to recommend similar type of an outfit for user. In our case, we mainly put attention to the outfit itself to see if the combination of kinds of clothing items is good or bad, rather than consider about the characteristics of the users or the scene behind the users, which makes our system so general that it can be used in every domain. What's more, our aim is to use the system to recommend good outfit to users based on their clothing items. Following figure show the outfit items in dataset.

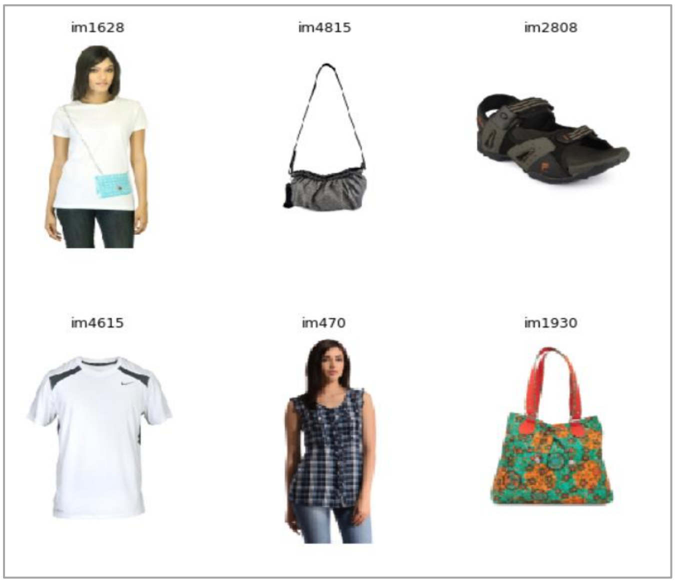

Figure 1. Outfit (Items) in dataset

The paper is organized as follows: Section 2 presents related work for this paper. Section 3 presents the proposed implementation. Section 4 explains the datasets used for our work and the experiments and results are presented. In section 5 conclusion of proposed work is given.

II. RELATED WORK

There are lots of methods achieved in fashion recommendation. For instance, customer ratings and clothing are utilized as considerations for fashion recommendation [6]. Similarly, user's personal preference and the history of clothing items have been tried [7]. Furthermore, some scholars found that the past statistics of clothes and accessories and current weather conditions as well as special occasions can provide a relevant recommendation on garment [8]. In order to meet different needs, an intelligent clothing recommendation system based on the principles of wearing fashion and aesthetic is studied [9]. In addition to the above work, Iwata et al. (2011) [10] offered a recommender system, utilizing fashion magazines' full-body photographs. In the same way, Sha et al. (2016) [11] extracted multiple features from images to analyze their contents in different attributes, such as fabric pattern, collar, and sleeve. Some garment system integrates the fashion themes and shapes professional designers' knowledge and perception to help them choose the most relevant garment design scheme for a specific customer [12].

A lot of work that has been done in various aspects of clothing recommendations. Works [13] - [18] all deal with recommendations. Earlier works like [17] and [18] made lexical recommendations. Lexical methods like Multimedia Web Ontology Language, Open Mind Common Sense and contextual knowledge have been used in these works in order to make recommendations. All these methods use some form of textual manipulation. In some of the later works [13] - [16], recommendations are visual-based. Images are analyzed instead of textual manipulation, in these works. Recommendations improved in these works because a lot more information is obtained from images. Earlier works like [9] - [11] find clothes that complement one another. This involves recognition of clothes in the images. Various works like [12], [24] have worked on clothing recognition. While [24] deal with image parsing, that is, finding the different components within the image. The works [12] - [14] have worked on clothing recommendations. However, not a lot of research has been done in recommending clothes based on the events (a party, a wedding, a meetup, a red carpet) at which they will be worn. If suggestions, consider such events then users will be able to look for clothes specific to their needs and will be inclined to buy the clothes recommended. The works [10] and [17] make scenario-oriented recommendations. However, [17] uses a text-based methodology. On the other hand, [10] uses Support Vector Machines on images to make recommendations. Of the other works in recommendations, [20] analyses personal style and [16] deals with clothing analysis based on the location of countries. Personalizing suggestions in this way has been known to increase the chances of a piece of clothing being purchased.

Recommendations can be improved if we analyze and find clothes which people pair together. This could be achieved by using deep neural networks. Using image processing to make recommendations has gained popularity over the last few years. Neural networks have made a lot of progress in image processing. Neural networks can be trained to identify specific features in images. But when similar features appear at different positions in the image, artificial neural networks cannot identify them [19]. Adding training images for all such images is not feasible. Instead, convolutional neural networks could be used, which identify features without bothering about their position. Works like [11], [15], [19] - [21] use convolutional neural networks to make clothing recommendations. Convolutional neural networks were made particularly popular by Krizhevsky, et al. [21]. They created a network known as AlexNet, which is considered as a turning point for neural networks. Convolutional Neural Networks (CNN) are a category of neural networks. A CNN usually has a convolutional layer along with a few other layers. The convolutional layer runs by applying the operation of convolution. An image could be represented as a matrix of its pixels. Consider Fig. 2 (a) to be our image matrix. And consider Fig. 2 (b) to be the filter. The convolution operation slides a window of the filter size over the image matrix and computes another matrix which is the convolved feature matrix (Fig. 2 (c))

$$\begin{array}{|c|c|c|c|c|}\hline 0&0&1&1&1\\\hline 0&0&0&1&1\\\hline 1&1&1&0&0\\\hline 1&0&1&1&0\\\hline 1&1&1&1&0\\\hline\end{array} \quad * \quad \begin{array}{|c|c|c|}\hline 1&0&1\\\hline 0&1&0\\\hline 1&0&1\\\hline\end{array} \quad = \quad \begin{array}{|c|c|c|}\hline 3&2&4\\\hline 3&3&2\\\hline 4&4&3\\\hline\end{array}$$

Fig 2 (a): Image Matrix Fig 2 (b): Filter Fig 2 (c): Convoluted Matrix

Recently, Siamese networks and triplet embeddings have also gained popularity [20], [22]. The main benefit of using Siamese networks is that they directly train for the problem at hand. For example, if we are trying to find similarity between two pieces of clothing, then Siamese networks will directly address this issue, instead of training for object detection and then finding the similarity. The research in [20] uses a Siamese CNN to find similar clothes. Triplet embeddings are beneficial because they help improve classification accuracy by extracting better features. The work [22] proposes a bidirectional cross-triplet embedding algorithm. They combine triplet embedding with cross-domain image retrieval. Scenario/event recognition is an important part of clothing recommendation in the proposed work. In the work [29], event recognition is performed using three steps: event concept discovery, training concept classifiers and prediction of concept scores. Tagged images are used to obtain concepts. Images of correlated concepts are clustered together. A feature vector is formed by taking a concatenation of all concept scores for an image. A classifier then classifies the image as a particular event. The work [30] performs event recognition in photo collections. They use a hidden Markov model for the same. However, this sort of event recognition does not make use of all the information available in the image. It uses tags associated with the image to find out objects within the image. Instead, we can detect objects within the image to identify the type of event. Using transfer learning has also gained popularity. Transfer learning is the process of using weights of an already trained model in order to train a new model on a similar task. In this way, we can avoid training an entirely new model from scratch and can instead concentrate on improving the task at hand. Transfer learning improves the time required to train a model significantly. In this paper, we make use of transfer learning to train models for object detection. We use Faster RCNN as the meta-architecture for object detection.

RCNN was the idea of Girshick et al. [3]. They came up with a good method to detect objects for the PASCAL VOC challenge, a popular object detection and classification challenge. This method drastically improved the accuracy of object detection as compared to its earlier works. Fig. 3 gives the system overview of RCNN. RCNN works by performing training on three separate models. The first model makes region proposals. These are regions with high possibility of containing an object. Each image has about 2,000 region proposals. RCNN creates region proposals using a method called Selective Search. RCNN will run CNN on each of the region proposals. So, when the region proposals are huge in number, it takes a long time for processing. That is why, RCNN requires about fifty seconds as average testing time per image. The second model trains Support Vector Machines (SVM) to classify which category the object belongs to. After this, the bounding boxes for the model generated are made more precise by training a linear regression model

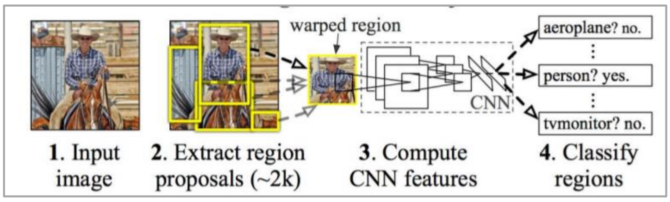

Figure 3: RCNN System Overview [3]

One of the authors of RCNN, R. Girshick, continued working on the challenges faced by RCNN and came up with a new algorithm, Fast RCNN [4]. Fast RCNN solves the issue of the slow RCNN architecture. In this new architecture, the region proposals are created in a similar manner to RCNN, by using Selective Search. However, CNN is run just once on the entire image instead of on all region proposals. Fast RCNN has a layer called the Region of Interest (RoI) pooling layer. In this layer, CNN features for every region proposed are obtained from the single CNN feature map. Then these features are pooled together. After this, instead of training an SVM model, a softmax layer is added to the model for classification. Also, a linear regression layer is added to the same model to get precise bounding boxes. Thus, just one CNN is run on the entire image. Because of this, Fast RCNN reduces the average testing time per image to around two seconds.

System Architecture

Following figure 4 shows the proposed system architecture of our fashion recommendation system.

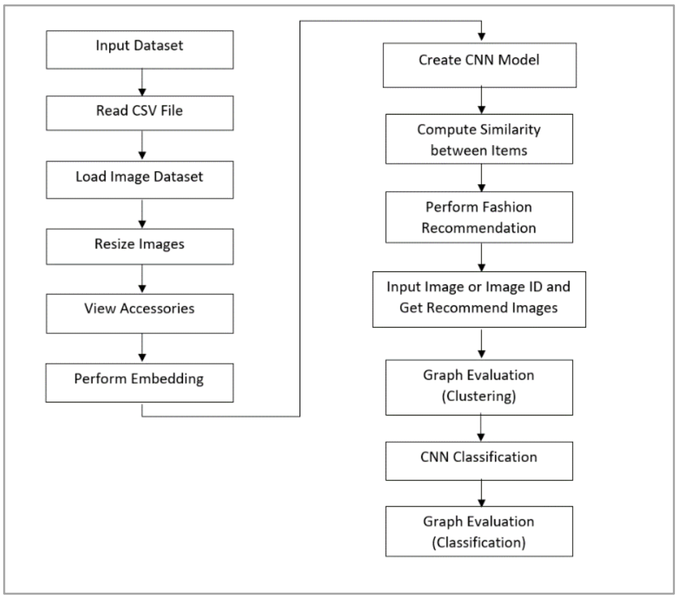

Figure 4. Proposed System Architecture

Following are different modules used in our proposed system architecture

1. Input Dataset:

For Fashion outfit recommendation we used kaggle dataset. Which consist of various categories as shown in Figure 5. The dataset is a collection of 14 GB images which are fully parsed and also it contains csv file. This dataset was collected from https://www.kaggle.com/paramaggarwal/fashion-product-images-dataset . This dataset has fifteen plus clothing and outfit categories like dress, jeans boots, etc.

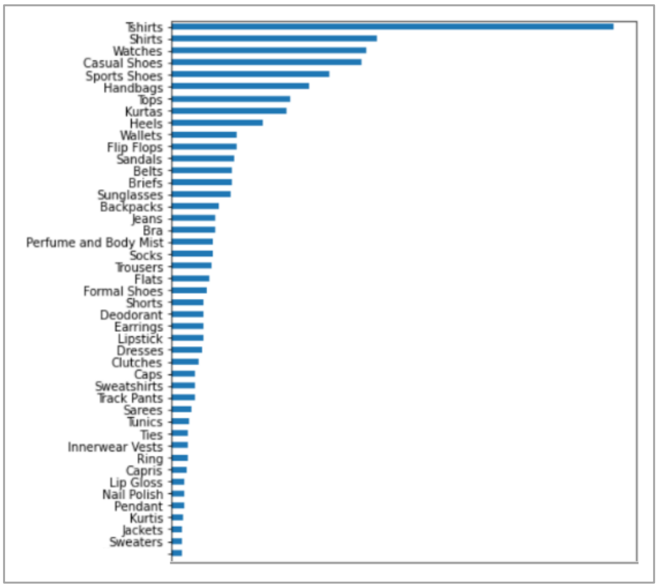

Figure 5. Fashion Outfit Categories

2. Read CSV file and Load Images

After loading the csv and images into our system all data can be view and it added in several objects. The various attributes are present in csv file which are id, image-id, gender, master-catogery, etc as shown in Table 1. After loading images into system image resize is performed to improve the performance of system and to make all images of same size.

Table 1. Fashion dataset with different attributes

	id	gender	masterCategory	subCategory	articleType	baseColour
0	15970	Men	Apparel	Topwear	Shirts	Navy Blue
1	39386	Men	Apparel	Bottomwear	Jeans	Blue
2	59263	Women	Accessories	Watches	Watches	Silver

3. Embedding

We will use embedding to identify similar items, this information will be used to recommend similar content. An embedding is a relatively low-dimensional space into which we can translate high-dimensional vectors. Embedding make it easier to do machine learning on large inputs like sparse vectors representing words. Ideally, an embedding captures some of the semantics of the input by placing semantically similar inputs close together in the embedding space. An embedding can be learned and reused across models. So a natural language modelling technique like Word Embedding is used to map words or phrases from a vocabulary to a corresponding vector of real numbers. As well as being amenable to processing by learning algorithms, this vector representation has two important and advantageous properties:

- **Dimensionality Reduction**—it is a more efficient representation
- **Contextual Similarity**—it is a more expressive representation

We are using the Embedding as input of the model, containing a reduced dimensionality but with much semantic information.

4. CNN Model Generation and Classification

Convolutional networks can be used to generate generic embeddings of any content. These embeddings can be used to identify similar items and in a recommendation process. Following are different steps included in CNN training.

Step 1: Choose a Dataset

For Classification we used pre-processes dataset of fashion recommendation

Step 2: Prepare Dataset for Training

Preparing our dataset for training will involve assigning paths and creating categories(labels), resizing our images.

Step 3: Create Training Data

Training is an array that will contain image pixel values and the index at which the image in the categories list.

Step 4: Shuffle the Dataset

Step 5: Assigning Labels and Features

This shape of both the lists will be used in Classification using the Neural Networks.

Step 6: Normalising data values and converting labels to categorical data

Step 7: Split Train and Test for use in CNN

Step 8: Define, compile and train the CNN Model

Step 9: Accuracy and Loss Calculation of model

5. Clustering / Compute Similarity of Items

Cosine similarity measures the similarity between two vectors of an inner product space. **Cosine similarity** is a measure of similarity that can be used to compare documents or, say, give a ranking of documents with respect to a given vector of query words. Let x and y be two vectors for comparison. Using the cosine measure as a similarity function, we have

$$sim(A,B) = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

where ||x|| is the Euclidean norm of vector x=(x1,x2,…,xp), defined as x12+x22+⋯+xp2. Conceptually, it is the length of the vector. Similarly, ||y|| is the Euclidean norm of vector y. The measure computes the cosine of the angle between vectors x and y.

6. Recommend similar fashion outfit

Once we find the top categories of clothes, we obtain items of clothing / outfit from the various images. These items of clothing have been worn at the events. The assumption is that if similar clothes are recommended to the users, they will be appropriate. Hence, we find out similar clothes using the CNN predict. We can pass image id or image as input to CNN predict algorithm, in return we get similar outfit or products as shown in Figure 6 and Figure 7. Figure 6 shows recommended shirts while Figure 7 shows recommended shoes.

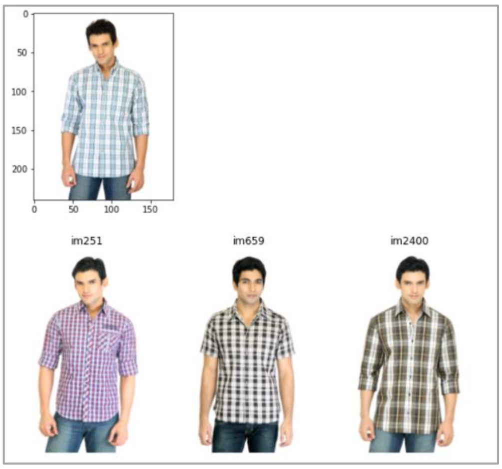

Figure 6. Input shirt image and 3 recommended shirt images

Figure 7. Input shoes image and 3 recommended shoes images

III. Results and Discussion

A. Experimental Setup

All experiments are performed on the Windows 10 platform. It is carried out on a device of 8 GB of RAM and a 256 GB SSD. For implementation, Python 3.7 Technology and Jupiter Notebook Tool was used to write the code for the models used.

B. Results

Figure 8 and Figure 9 shows the clustering results of master categories and sub- categories of cloths and outfit.

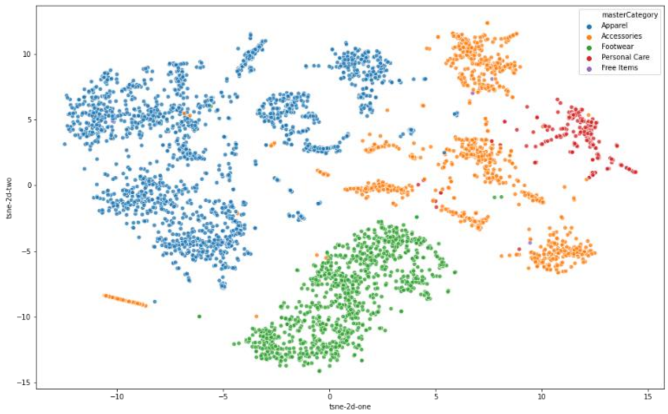

Figure 8. Master Categories of Outfit (Cluster)

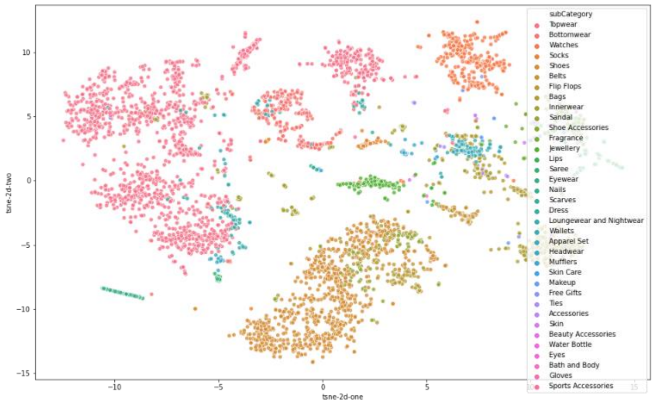

Figure 9. Sub-Categories of Outfit (Cluster)

Figure 10 shows the training and validation accuracy of CNN algorithm and Figure 11 shows the training and validation loss of CNN, with increasing number of epoch accuracy gets increased while loss gets decrease. Table 2 shows graphs reading of accuracy and loss.

Table 2: Training Accuracy and Loss with respect to number of epoch

Number of Epoch	Training Accuracy	Training Loss
2	0.831	0.115
4	0.838	0.084
6	0.842	0.076
8	0.848	0.071
10	0.844	0.065

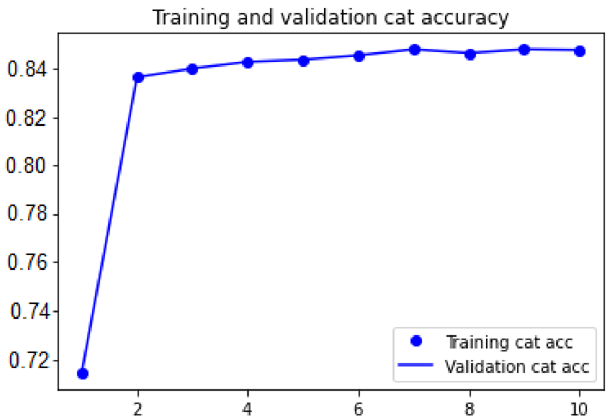

Figure 10. Training and Validation Accuracy Comparison of CNN

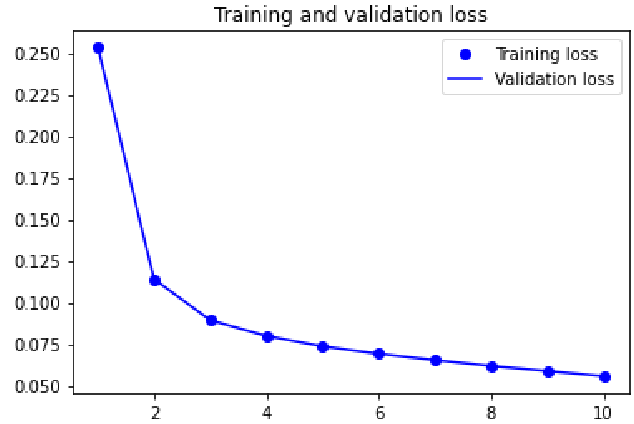

Figure 11. Training and Validation Loss Comparison of CNN

IV. CONCLUSION

Recommending clothes using images has made tremendous progress over the years. Ecommerce websites are hugely benefitted by this. As research in this field continues, more and more interesting methods have come to light. Work once started using text-based methods, turned to visual methods with image processing and use of neural networks, convolutional neural networks and now transfer learning with deep neural networks. Thus, we know that there is a common theme in the recent research carried out in the field of fashion outfit recommendation. This theme is analyzing images, finding out features in the images and classifying pieces of clothing in the image. One can understand that this methodology works for most systems. Also, the existing scenario-based recommendations for clothes do not fully utilize the capability of deep neural networks. This paper has introduced an improved CNN based novel approach to recommending outfit and also we make use of clustering techniques (embedding) to generate categories of outfit. Our experimental results show the accuracy of 84 percentages. Our model can be used in the outfit recommendation system, which is given a pool of cloth items it can recommend the users the best outfit generated from the input items.

V. REFERENCES

[1]. F. Isinkaye, Y. Folajimi and B. Ojokoh, "Recommendation systems: principles, methods and evaluation", Egyptian Informatics J., vol. 16, no. 3, pp. 261-273, 2015. https://en.wikipedia.org/wiki/Artificial_neural_network, Accessed: 12-Apr-2018].

[2]. R. Girshick et al, "Rich feature hierarchies for accurate object detection and semantic segmentation," in 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014. DOI: 10.1109/CVPR.2014.81.

[3]. R. Girshick, "Fast R-CNN," in 2015 IEEE International Conference on Computer Vision (ICCV), 2015. DOI: 10.1109/ICCV.2015.169.

[4]. S. Ren et al, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, (6), pp. 1137- 1149, 2017. DOI: 10.1109/TPAMI.2016.2577031.

[5]. X. Hu, W. Zhu, and Q. Li, “HCRS, A hybrid clothes recommender system based on user,” in Proceedings of the Ratings And Product Features, pp. 270–274, 2014.

[6]. L. Yu-Chu, Y. Kawakita, E. Suzuki, and H. Ichikawa, “Personalized clothing-recommendation system based on a modified bayesian network,” in Proceedings of the IEEE/IPSJ 12th International Symposium on Applications and the Internet (SAINT '12), vol. 59, pp. 414–417, July 2012.

[7]. C. Limaksornkul, D. N. Nakorn, O. Rakmanee, and W. Viriyasitavat, “Smart closet: Statistical-based apparel recommendation system,” in Proceedings of the 2014 3rd ICT International Senior Project Conference (ICT-ISPC '14), pp. 155–158, March 2014.

[8]. S. Liu, T. V. Nguyen, J. Feng, M. Wang, and S. Yan, “Hi, magic closet, tell me what to wear !,” ACM Multimedia, vol. 43, pp. 1333-1334, 2012.

[9]. T. Iwata, S. Watanabe, and H. Sawada, “Fashion coordinates recommender system using photographs from fashion magazines,” in Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI '11), pp. 2262–2267, Catalonia, Spain, July 2011.

[10]. D. Sha, D. Wang, X. Zhou, S. Feng, Y. Zhang, and G. Yu, “An approach for clothing recommendation based on multiple image attributes,” Web-Age Information Management, 2016.

[11]. L. C. Wang, X. Y. Zeng, L. Koehl, and Y. Chen, “Intelligent fashion recommender system: Fuzzy logic in personalized garment design,” IEEE Transactions on Human-Machine Systems, vol. 45, no. 1, pp. 95–109, 2015.

[12]. V. Jagadeesh, R. Piramuthu, A. Bhardwaj, W. Di and N. Sundaresan, "Large scale visual recommendations from street fashion images", Proc. of the 20th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data mining - KDD '14, 2014.

[13]. Y. Hu, X. Yi and L. Davis, "Collaborative Fashion Recommendation: A Functional Tensor Factorization Approach", Proc. of the 23rd ACM Int. Conf. on Multimedia - MM '15, 2015.

[14]. X. Zhang, et al., "Trip outfits advisor: location-oriented clothing recommendation", IEEE Trans. on Multimedia, pp. 1-1, 2017.

[15]. E. Simo-Serra, S. Fidler, F. Moreno-Noguer and R. Urtasun, "Neuroaesthetics in fashion: modeling the perception of fashionability", 2015 IEEE Conf. on Comput. Vision and Pattern Recognition (CVPR), 2015.

[16]. E. Shen, H. Lieberman and F. Lam, "What am I gonna wear? scenario-oriented recommendation", Proc. of the 12th Int. Conf. on Intelligent user interfaces - IUI '07, 2007.

[17]. L. Yu-Chu, Y. Kawakita, E. Suzuki and H. Ichikawa, "Personalized clothing recommendation system based on a modified bayesian network", 2012 IEEE/IPSJ 12th Int. Symposium on Applications and the Internet, 2012.

[18]. S. G. Eshwar, G. G. Prabhu J, A. V. Rishikesh, A. N. Charan and V. Umadevi, "Apparel classification using convolutional neural networks", 2016 Int. Conf. on ICT in Business Industry & Government (ICTBIG), 2016

[19]. A. Veit, B. Kovacs, S. Bell, J. McAuley, K. Bala and S. Belongie, "Learning visual clothing style with heterogeneous dyadic co-occurrences", 2015 IEEE Int. Conf. on Comput. Vision (ICCV), 2015.

[20]. A. Krizhevsky, I. Sutskever and G. Hinton, "ImageNet classification with deep convolutional neural

networks", Communications of the ACM, vol. 60, no. 6, pp. 84-90, 2017.

[21]. S. Jiang, Y. Wu and Y. Fu, "Deep bi-directional cross-triplet embedding for cross-domain clothing retrieval", Proc. of the 2016 ACM on Multimedia Conf. - MM '16, 2016.

[22]. Z. Liu, P. Luo, S. Qiu, X. Wang and X. Tang, "Deep Fashion: powering robust clothes recognition and retrieval with rich annotations", 2016 IEEE Conf. on Comput. Vision and Pattern Recognition (CVPR), 2016.

[23]. K. Yamaguchi, M. Kiapour and T. Berg, "Paper doll parsing: retrieving similar styles to parse clothing items", 2013 IEEE Int. Conf. on Comput. Vision, 2013.

Cite this article as :

Bhagyshree Pravin Bhure, Pratiksha Tulshiram Bansod, Monali Shivram Amgaokar, Savita Pralhad Lodiwale, Anjali Pravin Orkey, Ashish Mohod, "Fashion Outfit Recommendation Based on Deep Learning Model", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 8 Issue 3, pp. 216-225, May-June 2021. Journal URL : https://ijsrst.com/IJSRST218370