PRODUCT RECOMMENDATION USING OBJECT DETECTION FROM VIDEO, BASED ON FACIAL EMOTIONS

Kshitiz Badola[1], Ajay Joshi[2] and Deepesh Sengar[3]

[1]Department of Computer Science, Mahavir Swami Institute of Technology, Guru Gobind Singh Indraprastha University, New Delhi, India
[2]Department of Electronics, Deen Dayal Upadhyaya College (University of Delhi), New Delhi, India
[3]Department of Computer Science, Riga Technical University, Riga, Latvia

ABSTRACT

In today's world, with the increasing demand of products and their growing productivity from producers, customers sometimes failed to decide whether they are interested in buying a particular product or not. So author, here proposed a framework which deals with the buying of only items of interest, for a consumer. In our feature-set, whenever any consumer tends to watch any video from YouTube, it results in breakdown into several frames (frames per second), and from there we use object detection technique to detect each and every object in a particular frame, and then to find whether our consumer is interested in that particular object or not, we use facial emotion detector to check whether our user is happy, surprised, neutral or any other emotion. After viewing those products which are present in a frame of a video. Merging only those items of interest which were tend to fall for consumer's positive choices (emotions), we then used Amazon online marketing technique to recommend products selected by our framework.

KEYWORDS

Convolutional Neural Networks, Facial Expressions, Object Detection, ImageAI, Selenium, Machine Learning.

1. INTRODUCTION

Decades back when machine learning was not introduced, it was hard to predict any human being's interest, mood and ability. But with the growth of artificial intelligence and machine learning applications we can predict most of the things of any user like expressions, gestures, etc. Machine learning allowed us to train our data as well as our system according to what we want to predict from a user, which leads in performing of several tasks that would have otherwise required considerable human efforts to very much extent. Authors, in this paper provides a framework which will use few machine learning techniques, which will result in product recommendation only and if our consumer who is willing to buy product is interested or not in that product, this type of product recommendation is needed because many times consumer fails to decide his/her interest in buying a product. In this framework we use video object analysis to find objects in a particular watching video, by consumer. The chosen video by the consumer from YouTube has been divided into several frames with the help of our code work, 1 second for one frame (e.g. for 2 minutes video, 120 frames), then these divided frames are then used by object detection model to predict all kind of objects/products. From there we pull out maximum

David C. Wyld et al. (Eds): SIGI, CSTY, AI - 2020
pp. 51-56, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.102006

occurring (counted) objects only. Then using another technique of CNN (Convolutional Neural Network), emotion detector model is used to predict consumer interest (like happy, sad, neutral, surprise, fear, disgust and angry) and finally with those interest factors, one or more products are targeted for prediction in recommendation from Amazon online store. Our purpose for choosing this project was not only to make consumers meeting with products of their positive choices but also, to ensure the increasing productivity in the market for only those products/materials for which consumer's interest is more severe/maximum, this will be highly useful for producers too. And this will also help in better economic growth and better substantial development over worldwide.

2. TARGETING MILESTONES

We used object detection technique to detect each and every product in the video using every frame per second from that video watching by user at that particular time period. We maintained a counting system to select maximum occurring object in a particular video. To ensure the maximum accuracy of product we manually removed human beings from detection criteria so that detection could be done on products only. Facial emotion detector used in this project is using CNN [1]to classify of data into different labels[2] [3] to provide user's emotion time to time and categorise them into seven categories like happy, sad, angry, disgust, fear, neutral and surprise. And from those we use only happy, surprise and neutral emotions to find the interest of any consumers watching that video as these three emotions were usually common for a positive interest showed by a consumer in any object at the time of purchasing/buying. The purpose for choosing facial emotion detector for this product recommendation project was because many previous research papers used audio detection [4], speech to text detection [5], object detection [6], etc. in a video to find an object of interest, but in this paper authors tried different strategy to recognise the current mood (emotion) of consumer to decide his/her product of interest. At last after having the results from emotion detector linked with object detector for a particular object (one or more than one product) is then used in recommendation using Amazon online store directly splashing those items of interest into Amazon page so that our consumer can now easily select his/her interested product for ordering/buying that product. Non interested products will not be selected by our framework because they showed negative choice of interest of a user like sad, angry, disgust and fear, so we ignore such objects/products and don't open Amazon store link for such cases.

3. OBJECT DETECTION

For accessing object detection model we need an image and to make that model work in our project we needed several images or frames because we are targeting videos from YouTube. So to get those frames, we used Selenium library to access the YouTube platform, then we set accurate pixels of a particular scene from that video(to focus only on the video part), leaving the background of YouTube behind, undetected and not selected to be a part of a frame. So that, now we are able to focus on our product with more clarity and accuracy. After selecting pixels for frames of a particular scene from that video we designed our code to capture frames continuously (per second) till the time our work is terminated and store those frames in other working directory of our project. What if our consumer tries several videos of many different products? Our work will get every frame from every (as many) video he/she is watching for either one or more than one product.

Now we used object detection for predicting different products. Model here used is a retrained model, i.e. resnet50_coco_best_v2.0.1.h5 [7] to save our training time, also we used FAST-RCNN (Region based Convolutional Neural Network) [8] system in our project, so that we could

predict our products from video frames. We used imageAI library which will predict all the objects in a selected frame then return the objects with its respective coordinates. Also to improve our object detection model we omitted/ignored human beings detection from the frame, just to focus only on products. After getting all the objects by their name and their respective coordinates from the frame we find out the area of each object in that frame and built our model in a way such that objects with larger areas in a particular frame should be selected because consumer's interest will most probably lean towards the object that are with huge area in a particular frame of the video. After selecting single or multiple objects from a frame as a result, we tried to find the number of times a particular object has been displayed in the video and with their count we proceeded for further recommendation process. For that authors maintained a list where objects from all frames are present with their respective counts, and highest number of count will be considered as prediction. For example, if we are watching two videos (one after other), an unboxing of laptop and P.C. from YouTube, object detection will result in targeting monitor, keyboard, C.P.U., etc. from that video. For more clarity, let us suppose any user watched two videos of different topics, like one video of microwave and other video of phone, the result after running of our framework will be of objects like oven, microwave, charger, phone, screen, etc.

4. FACIAL EMOTION DETECTOR

Our next milestone was to predict emotions of a consumer while watching any video and for that authors used TensorFlow to train the model. Kaggle dataset [9] was used in our model which was trained by several competitors at the time of competition in year 2013. Competition itself was conducted by Kaggle only, after giving an overview for our model, we found 3500 datasets in which we trained 2800 successfully and rest dataset were used as testing. Authors predicted accuracy was 92.10% on 25 epoch with the validation loss of 21.96% and with knowledge of Deep Neural Networks (DNN) we used dataset named Fer2013 and we design our own neural network which is depicted in figure 1. Authors at this point, named this model as 25.h5, now this

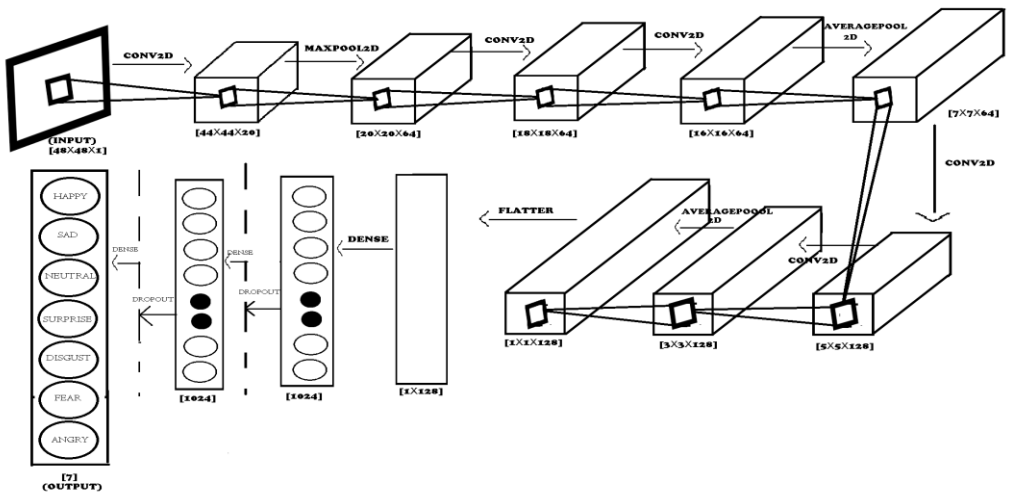

FIGURE 1. Model description

model will take input feeds and will predict output as classified between 7 categories that are happy, neutral, surprise, sad, angry, disgust and fear. Then according to our project we needed positive interest for a consumer so we mainly targeted in 3 emotions happy, neutral and surprise. Those 3 emotions will help in product recommendation as they are likely to be termed as

emotions for positive interest. If our user gets sad, angry, disgust or fear types of emotion while watching any video, it clearly results in his/ her negligible interest in that video so we will not provide product recommendation for such cases, this part is also important, because for good product recommendation system, all kinds of needs and emotion of a consumer should be judged properly and wisely. This model will show updated results time to time after executing it, till the time of termination so that our users need for each product could be refreshed at all time. And list of interested products could be maintained for product recommendation from Amazon online store.

5. MERGING RESULTS AND PRODUCT RECOMMENDATION

For merging results of selected product from frames and decided emotion, we used outputlinks.py. This will lead in reading of file and if our user is looking interested (happy, surprise or neutral) in any object from a frame of a watching video section, it will present a URL (Uniform Resource Locator) which will consist of a product that we obtained from object detection model based on our positive emotions (from emotion detector) linked with Amazon online store. So that anyone, anywhere in the world could access this product recommendation system for his/her interest directly with the product link available in Amazon store (if product not available at Amazon, then no recommendation). We are using selenium to access Amazon store from a chrome browser and from there our consumer can directly buy his/her product. What if we land in a position where we are interested in two or more products in a video? For that our work will present as many links, and open those links automatically on new tabs, and generate those URLs (as discussed) for only of products (one or more) that comes under consumer's positive interests according to our framework.

6. RESULTS AND DISCUSSIONS

Thus to ensure our working project based on user's individual product preferences, complex connections have been made and we surveyed 30 people (consumers) for our project and out of them we were getting positive results from 28 consumers for our product recommendation from Amazon store successfully only with the products of positive interest from a consumer. Our project for the recommendation of interested products using consumer choice is now successful. This framework if, used by any company/organization can lead in their better products advertisement as they will be targeting for consumer's product of interest. This framework can also be used in creating different types of advertisement content as they can trace the latest interest of the audience and their product preferences. For our working framework, we also present a block diagram of complete model which is depicted by figure 2.

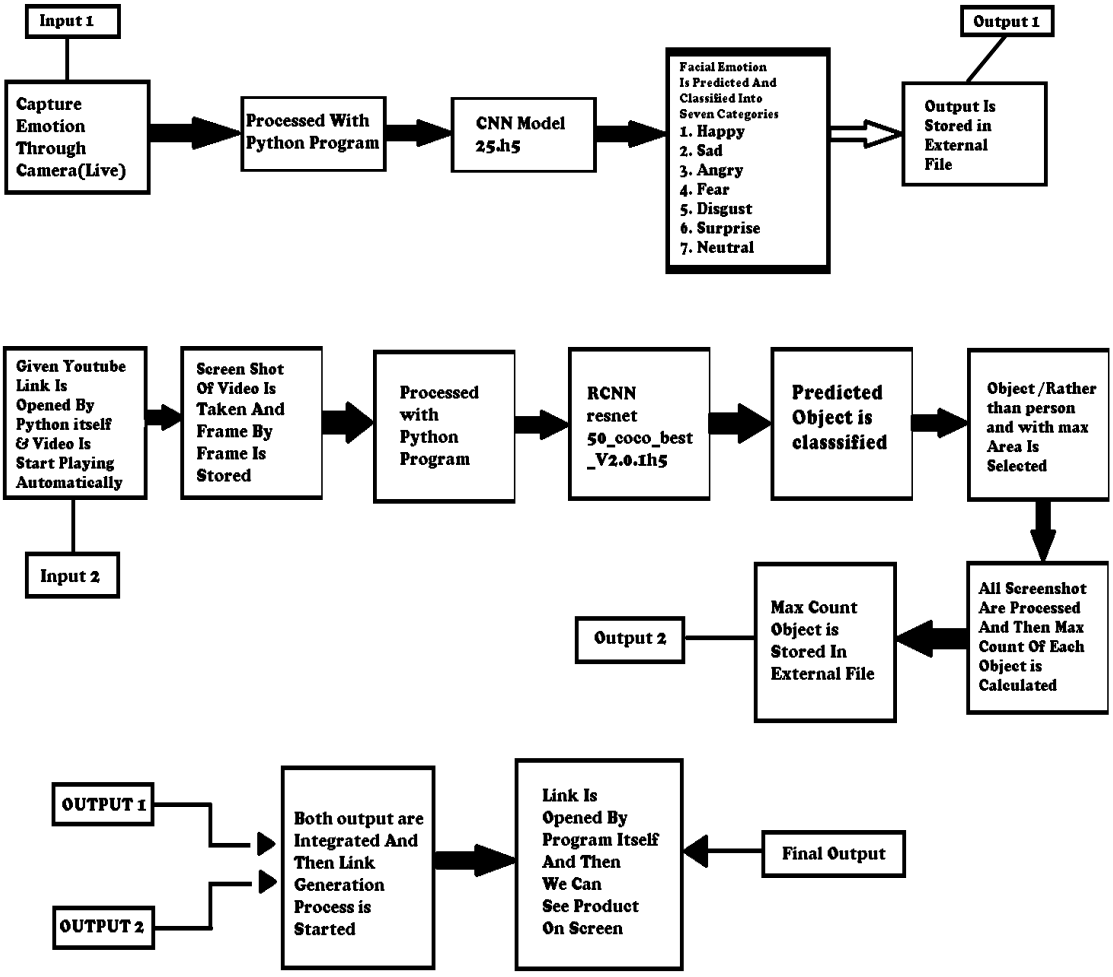

FIGURE 2. Final block diagram for working of project

7. CONCLUSIONS AND LIMITATIONS

As for now, our project is working successfully in recommending products of consumer's choice using their facial emotions and object detection, from any YouTube video. For the future work, we would like to combine all the feature-set and make it work for the real-time database system, not only YouTube video but also for other platforms like Facebook and Instagram posts. Also we would like to improve accuracy in our project by adding more classified emotions. We can make our own database for product recommendation where we can track each and every updated record of products of interest. Also we would to like to add more features in this for capturing reactions of consumers for products like audio detection, speech to text detection and any other machine learning technique which will lead in improvising our project accuracy

.

With our working project, there are some limitations that can be faced by any user at time of working on our framework, that are, our user who is watching video must be watching a product based content (for detection of product), otherwise no product will be selected for product recommendation for e.g. if any user watches fighting scene in a video, our framework will fail to detect any product from that scene. Another drawback could be, the API for our project is not

ready yet, so it will be difficult to run this work in any android system. There is one more minimal drawback which could affect our project, i.e. only a good processor P.C. would be highly appreciated for running our complete project, otherwise there is a chance of lagging if run in a slow processor system.

REFERENCES

[1] Yu, L., Li, B. and Jiao, B., (2019). Research and Implementation of CNN Based on TensorFlow. *IOP Conference Series: Materials Science and Engineering*, 490, p.042022. Available at: https://iopscience.iop.org/article/10.1088/1757-899X/490/4/042022.

[2] Ertam, F. and Aydin, G., (2017). Data classification with deep learning using Tensorflow. *2017 International Conference on Computer Science and Engineering (UBMK)*, Available at: https://ieeexplore.ieee.org/document/8093521
.

[3] Shiddieqy, H., Hariadi, F. and Adiono, T., (2017). Implementation of deep-learning based image classification on single board computer. *2017 International Symposium on Electronics and Smart Devices (ISESD)*, [online] Available at: https://ieeexplore.ieee.org/document/8253319.

[4] Smadi, T., Al Issa, H., Trad, E. and Smadi, K., (2015). Artificial Intelligence for Speech Recognition Based on Neural Networks. *Journal of Signal and Information Processing*, 06(02), pp.66-72. Available at: https://www.scirp.org/journal/paperinformation.aspx?paperid=55265.

[5] Nassif, A., Shahin, I., Attili, I., Azzeh, M. and Shaalan, K., (2019). Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access*, 7, pp.19143-19165. Available at:https://ieeexplore.ieee.org/document/8632885/.

[6] Galvez, R., Bandala, A., Dadios, E., Vicerra, R. and Maningo, J., (2018). Object Detection Using Convolutional Neural Networks. *TENCON 2018 - 2018 IEEE Region 10 Conference*, Available at: https://ieeexplore.ieee.org/document/8650517.

[7] GitHub.2020. Release Models For Image Recognition And ObjectDetection • Olafenwamoses/Imageai. Available at: https://github.com/OlafenwaMoses/ImageAI/releases/tag/1.0/.

[8] Girshick, R., (2015). Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, [online] Available at: https://ieeexplore.ieee.org/document/7410526.

[9] Verma, R., (2020). *Fer2013*. Kaggle.com. Available at: https://www.kaggle.com/deadskull7/fer2013.

AUTHORS

Mr. Kshitiz Badola pursuing B.Tech, Computer Science course from Guru Gobind Singh Indraprastha University, India. He had applied various A.I. tools in developing android systems, also he has immense knowledge in Deep Neural Network and Natural Language Processing. He is an assistant teacher of computer science department of his university.

Mr. Ajay Joshi graduated (by 2020) B.Sc. (hons) Electronics from Deen Dayal Upadhyaya College, University of Delhi, New-Delhi, India.

Mr. Deepesh Sengar is pursuing B.E. Computer Systems course from Riga Technical University, Latvia. He applied various programing tools and made the algorithm more efficient. He is the founder and lead developer of Quiqle Corporation and a tech enthusiast.