Content based advertisement recommendation system

Milon Bhattacharya
Indian Institute of Technology Ropar
milon.20csz0017@iitrpr.ac.in

***Abstract*—The aim of this study is to develop tools and techniques which can be used to recommend content-relevant advertisements. Current approaches are mainly based on a collaborative approach, where based on the previous browsing history of the customer, we display/show products that are either similar or even the same. This idea, simple in its approach and implementation has multiple problems namely, lack of relevance, concerns of privacy, and a strong bias towards the training data. A content-based approach, on the other hand, would only focus on the multimedia object (video, music, subtitles, etc) itself while trying relevant items. This would remove the dependency on the viewer and would be agnostic of the time, place, and person the content is being displayed. Also, the relevance of the ad is solely dependent on the content and hence does not need to be computed multiple times, thus reducing computation efforts. This is because, once a list of items has been decided as a recommendation, they can be saved as metadata and displayed for all future occurrences.**

***Index Terms*—CBIR, Object Detection, Fashion**

I. Problem Definition

For the purpose of this study, the multimedia object being considered is video (keeping in mind video streaming services). Further, the items to be displayed are assumed to be are clothing items from an e-commerce player. Each such item would consist of a set of images and a description. The aim would be to recommend structurally similar items from the catalog based on what is there within the video. All such previous attempts [8], have focused on a single modality, either image, audio, or text. The proposed approach would use video and image to make a holistic recommendation. As an outcome of the study, we aim to produce a tool, which would demo the above-mentioned functionality. Also, the perceptive quality of the recommendations would be recorded using a custom-made tool. This would look at measuring the quality of the recommendations, whether they are relevant to the video being displayed. The model will be tested on publicly available datasets. Namely, Deepfashion2 and iMaterialist. Also, to benchmark the recommendations we plan to use the movingfashion dataset.

II. Literature Review

In [2] The authors created a video to recommended image dataset which can be used to train models. Also, they developed a base recommendation model called SEAM RCNN to learn such recommendations. For the study in fashionpedia [5] Instance segmentation and fine-grained attribute categorization are performed in a single method. The attributes were taken from a custom ontology. The segmentation was performed using attribute mask-RCNN. Finally, a new evaluation metric was proposed. The complete study was done by re-labelling Deepfashion2 and Imaterialist datasets. Another study [7] *Buy-the-look* recommended items from the catalog using the following steps, using the query image, localize the human using pose estimation, within the pose find all fashion items (shirts, belt trouser, etc). Using the pose estimate classify as the front, backside, etc. The fashion items are detected using an ANN trained on Imaterialist dataset, which are extracted as bounding boxes. Misclassified instances are manually labeled and the model is re-trained. This ensure continues improvement. Using a tripletnet embedding find the most similar image from an existing catalog and display the n most relevant items. The catalog items have been arranged using different classes (same as the ones from IMAT). The search only happens within the indices of the specific categories.

III. Proposed Solution

1) *Step 1*: Using the highest possible resolution of the input video, perform scene detection. Select f frames from each of the detected scenes to collect a total of F frames from the complete video.
2) *Step 2*: On each of the frames, apply a pose detection to find all possible human figures. Crop out each of the humans.
3) *Step 3*: Apply object detection on each of the areas where there are humans. Select out the detected fashion objects.
4) *Step 4*: Find embedding of the cropped images.
5) *Step 5*: Using the embeddings of the catalogue items, find the most similar item from the cropped image.

IV. Implementation

1) For the first step, a maskRCNN [4] model is used for detecting the fashion objects in an image. The model is an enhancement over the Faster RCNN, where it predicts a mask for each of the proposals over and above the category and bounding box.

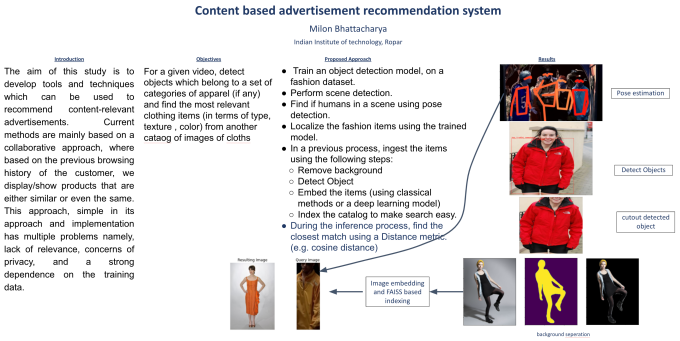

Fig. 1. Different stages in the pipeline.

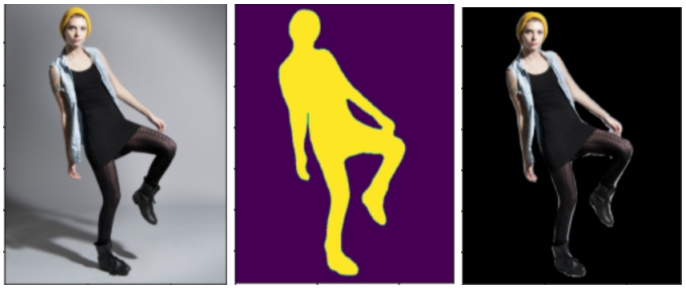

Fig. 2. Different stages in the pipeline.

2) The pretrained model is modified to change the number of classes in the output layer, which is 47. This is the number of classes in the iMaterialist dataset [3].
3) The dataset has approx, 50,000 images in which different types of clothing have been identified and labelled (e.g. shirts, trousers etc.) in MSCOCO format [6].
4) To begin with, the background from each of the catalogue images is removed using a Resnet-101. This is a pre-trained model on the PASCAL Visual Object Classes Challenge 2012 dataset.[1]
5) Once the training is completed, we can use the model to generate garment categories and bounding boxes from unknown images.
6) The previous step detects garment types and weakly labels them. These detected garments are cropped out and grouped according to the garment category.
7) As a next step, we take the video where recommendations have to be made.
8) Scene detection is applied, which divides the video into discrete scenes, random frames from each scene are selected (with an overall limit on the total number of frames)
9) One each frame, we apply pose detection to detect the presence of human subjects. This would help us weed out non-relevant video/frames.
10) On each of the detected humans, we apply fashion object detection and cutout the garments.
11) Finally we embed all the images to be searched and the catalog images as a fixed length vector.
12) The vectors from the catalog images are indexed using a library called FAISS. Such a hashing based search mechanism ensures almost a constant search time, irrespective of the search space.

Fig. 3. Random frame with detected humans (https://www.youtube.com/watch?v=6M6samPEMpM&ab_channel=BackstreetBoysVEVO)

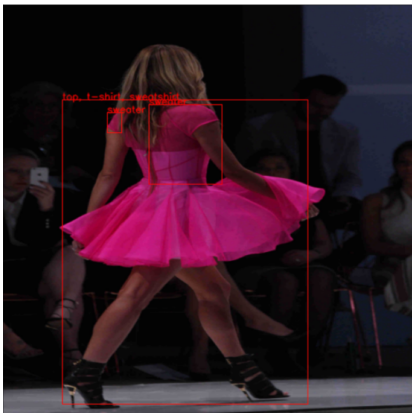

Fig. 4. Detection of fashion objects

V. Observations

VI. Conclusion and Future Work

1) One of the biggest challenges is measuring the quality of of recommendations. Object similarity for clothing items differs from person-to-person. Someone may perceive similarity based on color, other may consider the style (collar versus round-neck) as a measure of similarity.

Fig. 5. A sample matching result

2) Another problem faced was that the images were taken in different poses and with varied background, lighting conditions. This makes making recommendations even more difficult.
3) The dataset which is openly available may may not have similar garment items in the first place, hence any recommendations will be irrelevant.
4) As, future work, an attempt will be made to use the Movingfashion dataset [2] to benchmark the approach.
5) Identify main character
6) Seek user feedback if recommendations are good

REFERENCES

[1] Mark Everingham and John Winn. "The pascal visual object classes challenge 2012 (voc2012) development kit". In: *Pattern Analysis, Statistical Modelling and Computational Learning, Tech. Rep* 8 (2011), p. 5.

[2] Marco Godi et al. "MovingFashion: a Benchmark for the Video-to-Shop Challenge". In: *arXiv preprint arXiv:2110.02627* (2021).

[3] Sheng Guo et al. "The imaterialist fashion attribute dataset". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*. 2019, pp. 0–0.

[4] Kaiming He et al. "Mask r-cnn". In: *Proceedings of the IEEE international conference on computer vision*. 2017, pp. 2961–2969.

[5] Menglin Jia et al. "Fashionpedia: Ontology, segmentation, and an attribute localization dataset". In: *European conference on computer vision*. Springer. 2020, pp. 316–332.

[6] Tsung-Yi Lin et al. "Microsoft coco: Common objects in context". In: *European conference on computer vision*. Springer. 2014, pp. 740–755.

[7] Abhinav Ravi et al. "Buy me that look: An approach for recommending similar fashion products". In: *2021 IEEE 4th International Conference on Multimedia Information Processing and Retrieval (MIPR)*. IEEE. 2021, pp. 97–103.

[8] Jong-Hyuk Roh and Seunghun Jin. "Personalized advertisement recommendation system based on user profile in the smart phone". In: *2012 14th International Conference on Advanced Communication Technology (ICACT)*. IEEE. 2012, pp. 1300–1303.