CS682 project proposal: Fashion recommendation via neural re-ranking

Vishal G
UMass Amherst
vishalg@umass.edu

Srinivas
UMass Amherst
<insert_here>@umass.edu

Harshitha
UMass Amherst
<insert_here>@umass.edu

Abstract

In our course project for CS682, we propose to investigate how we can improve fashion recommendation for e-commerce websites with a neural network. We propose to improve the existing retrieval pipelines via training a neural network with custom extracted features, improve personalization of retrieved items and come up with new evaluation measures to evaluate the shortcomings of existing measures.

1. Introduction

Image retrieval is a fundamental task in information retrieval, where given some text about the image description, we wish to retrieve top-k images efficiently. In the course of our project we wish to explore how one can improve the relevance of the image retrieval system given some descriptiont text for the task of fashion recommendation. We wish to train a neural networks for learning to rank to improve the relevance of top-k retrieved items via re-ranking by relevance using extracted custom features.

2. Previous works

There have been numerous works in the field of image retrieval and learning joint image-text join representations which can be used to retrieve text given image and vice-versa (via maximum inner product search). [8] is a foundational work on that which learns the representations of images and text jointly via contrastive learning. Once a top-k item list is retrieved to optimize recall, a second step of re-ranking is often performed to optimize precision [9].

Furthermore, recent works also have explored the relevance of personalization while recommendation items to a user based on a user profile [10]. Works such as [4] explore on personalization and interpretability of fashion recommender systems.

3. Datasets

For the task of retrieving items (images) given text, there are many datasets with aligned text and images with the most popular one being MSCOCO [2]. Other datasets such as Flickr30K [5], Oxford5K [7], Google landmarksv2 [11], Deep fashion [3] and INRIA holidays [6]. For our project on fashion recommendation we plan to use Fashion 30K [1] and [3] to benchmark our pipeline and use metrics such as precisionin top-k results (P@k) and mean average precision (MAP@k) to evaluation.

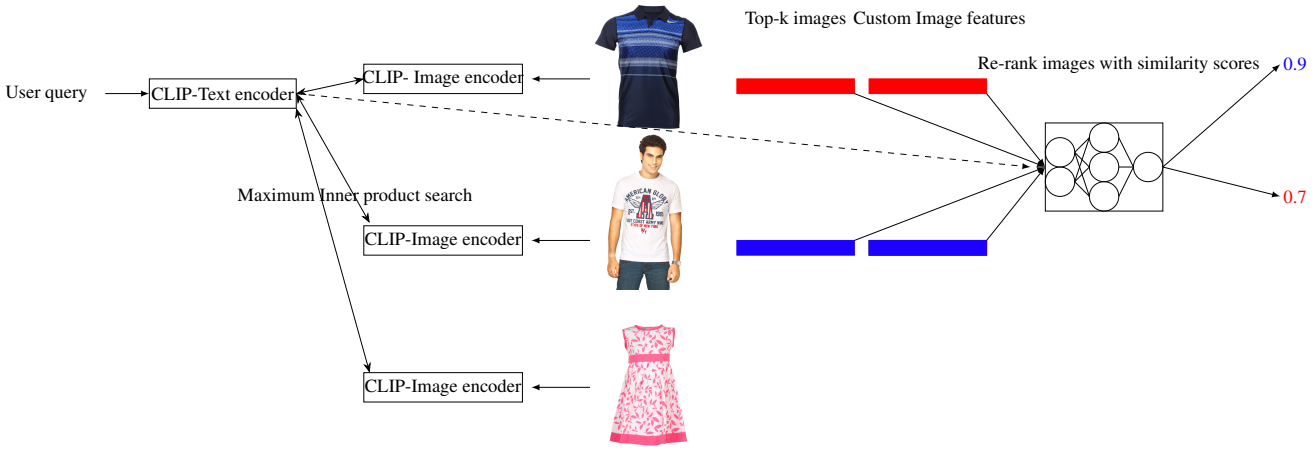
4. Our proposal

Many of the previous works, use CLiP embeddings of texts and images and compute maximum inner product search to retrieve images given text and vice-verse. We wish to explore a pipeline where, we re-rank the retrieved images, via learning a scoring similarity function to improve the P@k of the system. We propose to use features extracted from the given image such as dimensions of the bounding box, color of the item and so on and pass them to a neural network along with the input query to re-rank the top-k retrieved items.

We also wish to do the following extensions to this

1. **Rationale behind custom extracted features:** We intend to perform ablation for features such as size of the bounding box of the dress, width of the collar of the dress, etc from image segmentation models, and asses why they might be useful signals for re-ranking as they might indicate a good ”fit” between image and required product characteristics. **Harshitha to fill in details and add in the proposal and review all sections and image**
2. **Image generation and then retrieving:** Use Stable diffusion to generate images given text description and then use that generated image to retrieve similar images/evaluate images **Srinivas to fill in details and add in the proposal and review all sections and image**

Vishal to put in an image explaining our pipeline



References

[1] Param Aggarwal. Fashion product images dataset, 2019.

[2] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014.

[3] Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[4] Charles Packer, Julian McAuley, and Arnau Ramisa. Visually-aware personalized recommendation using interpretable image representations, 2018.

[5] Bryan A. Plummer, Liwei Wang, Chris M. Cervantes, Juan C. Caicedo, Julia Hockenmaier, and Svetlana Lazebnik. Flickr30k entities: Collecting region-to-phrase correspondences for richer image-to-sentence models. *CoRR*, abs/1505.04870, 2015.

[6] Adrian Popescu, Etienne Gadeski, and Hervé Le Borgne. Scalable domain adaptation of convolutional neural networks. *CoRR*, abs/1512.02013, 2015.

[7] Filip Radenovic, Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, and Ondrej Chum. Revisiting oxford and paris: Large-scale image retrieval benchmarking. *CoRR*, abs/1803.11285, 2018.

[8] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision, 2021.

[9] Xi SHEN, Yang Xiao, Shell Xu Hu, Othman Sbai, and Mathieu Aubry. Re-ranking for image retrieval and transductive few-shot classification. In *Advances in Neural Information Processing Systems*, pages 25932–25943. Curran Associates, Inc., 2021.

[10] Yuxin Tian, Shawn Newsam, and Kofi Boakye. Fashion image retrieval with text feedback by additive attention compositional learning. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 1011–1021, 2023.

[11] T. Weyand, A. Araujo, B. Cao, and J. Sim. Google Landmarks Dataset v2 - A Large-Scale Benchmark for Instance-Level Recognition and Retrieval. In *Proc. CVPR*, 2020.