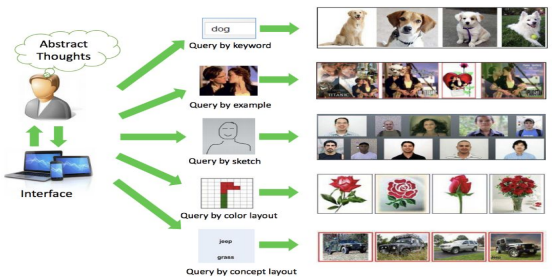
DeepSeek: Content Based Image Retrieval

Tanya Piplani (tanyapiplani@berkeley.edu)

Berkeley SCHOOL OF INFORMATION

Introduction



DeepSeek a natural language processing based deep learning model that allows users to enter a description of the kind of images that they want to search, and in response the system retrieves all the images that semantically and contextually relate to the query.

Data

Caption Generation Model

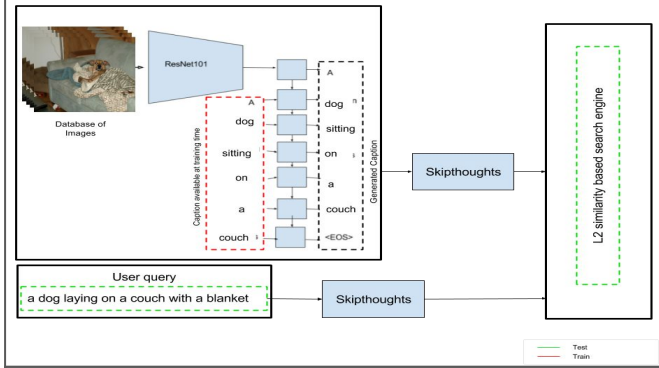
MS COCO 2015 dataset was used for training the caption generation model. 80k training images, 20k validation and 20k test images.

Image Retrieval Models

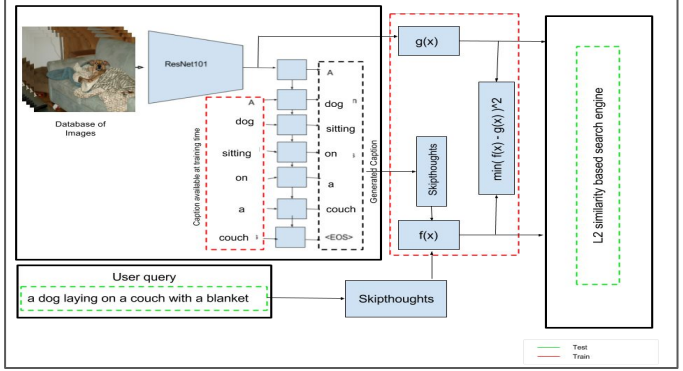
For training the Embedding based retrieval model, the generated captions on a subset of the train set of MS COCO 2015 along with its images were used (10k images + captions). The test set consisted of a subset of the test set of MS COCO (10k images + captions).

Models

Caption Based Retrieval



Embedding Based Retrieval



Quantitative Metrics

Caption Generation Model

Model	BLEU-1	METEOR	ROUGE-L	CIDEr-D
Our	0.828	0.280	0.603	0.692
SOTA	0.953	0.375	0.734	1.270

Image Retrieval Models

Model	Pr@1	Pr@3	Pr@5	Time
Caption Based	0.729	0.845	0.905	3.89 sec
Embedding Based	0.683	0.857	0.912	4.22 sec

Analysis & Conclusion

We see that both the caption based retrieval system and the embedding based retrieval system do a good job at content based image retrieval. The embedding system while slow, due to the need of calculation of embeddings at query time, is slightly more accurate when precision@5 is considered. More training and GPU optimization could make it faster and more accurate.