Latent Topic Modelling for Content Based Retrieval

By-
Aditi Singh(14IT104)
Parvathi M.H(14IT130)
Aparna PL(14IT132)

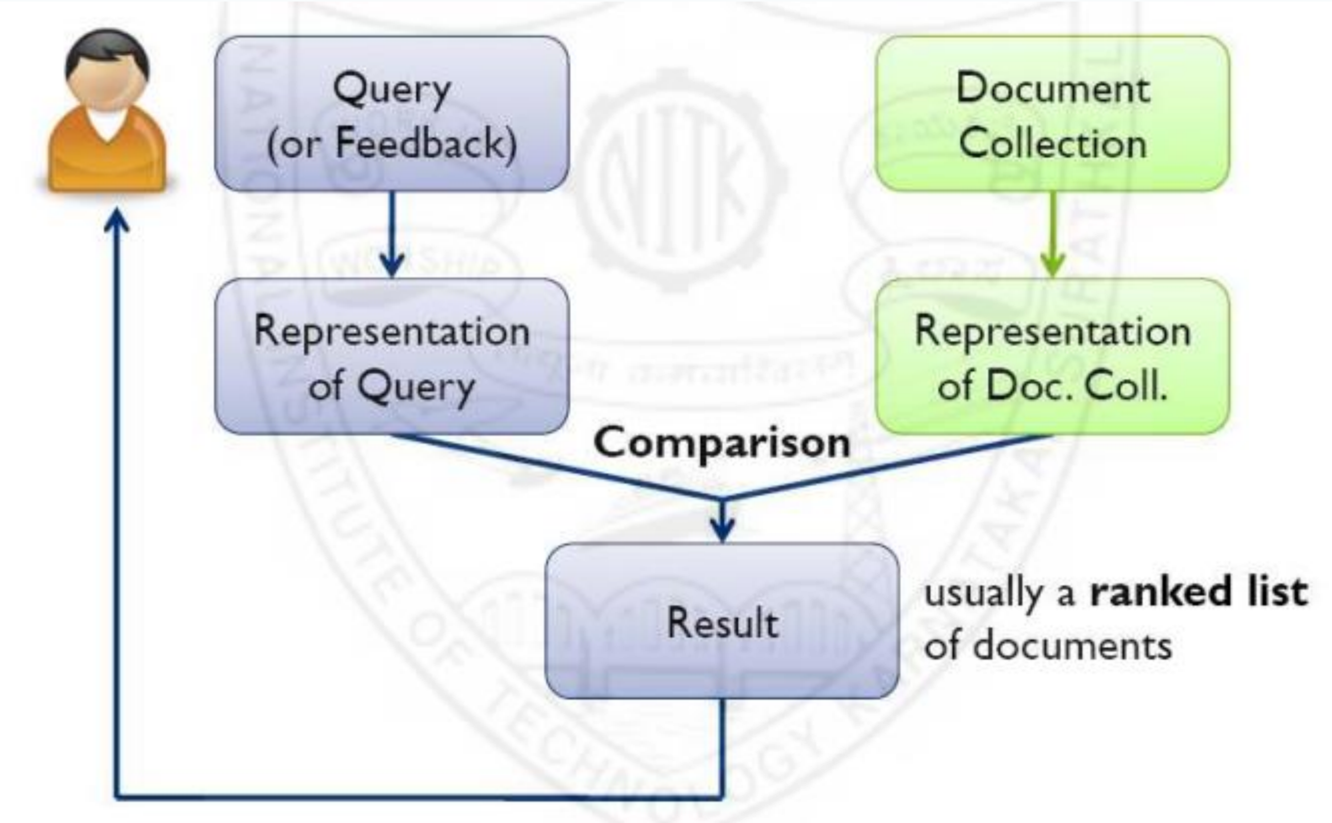
Introduction

- Content-Based Retrieval (CBR): providing users with images or videos which satisfy their queries.

- Similar images/videos found based on content of the query image/video (not from tags/metadata)

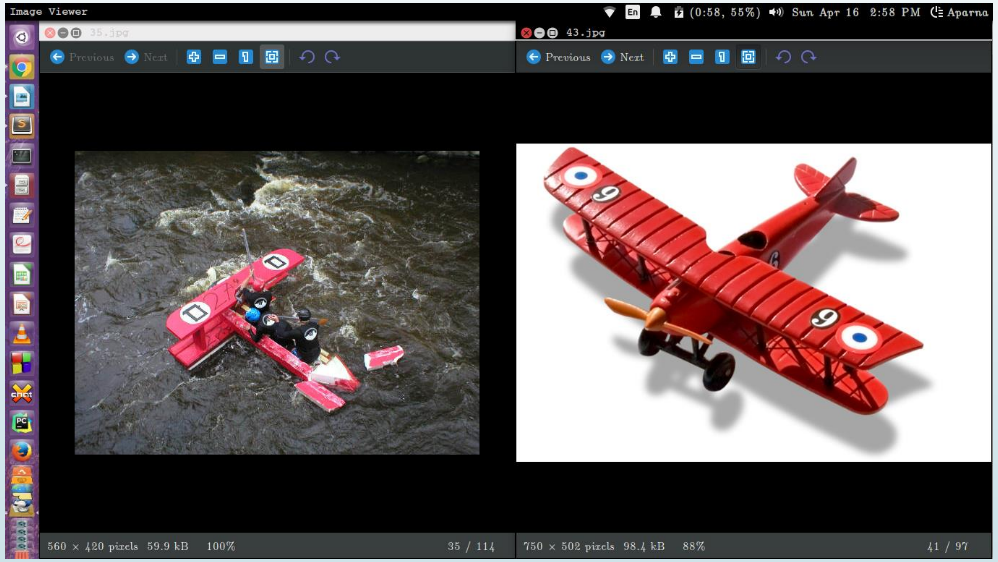
- Should satisfy information need that users have in their minds while searching for similar items from large collection of documents.

A Generic CBR system



Challenges of CBR

- Low-level features (such as shape, color etc.) is not enough for uniquely discriminating across different images/videos
- Semantic gap between low-level features and high-level concepts in images



- Class labels unknown, retrieval is performed in an unsupervised fashion

Problem statement

- To develop a novel Content-Based retrieval system which considers semantic (hidden)nature of the documents and not just visual features.

Related Work

- Probabilistic models with hidden/latent topic(LDA, LSA, pLSA) are popular in the document and language modeling community.

- Have been introduced and re-purposed for image content analysis tasks such as scene classification[1], object recognition[2] and image annotation[3].

Algorithms

- Bag of words
- Topic models

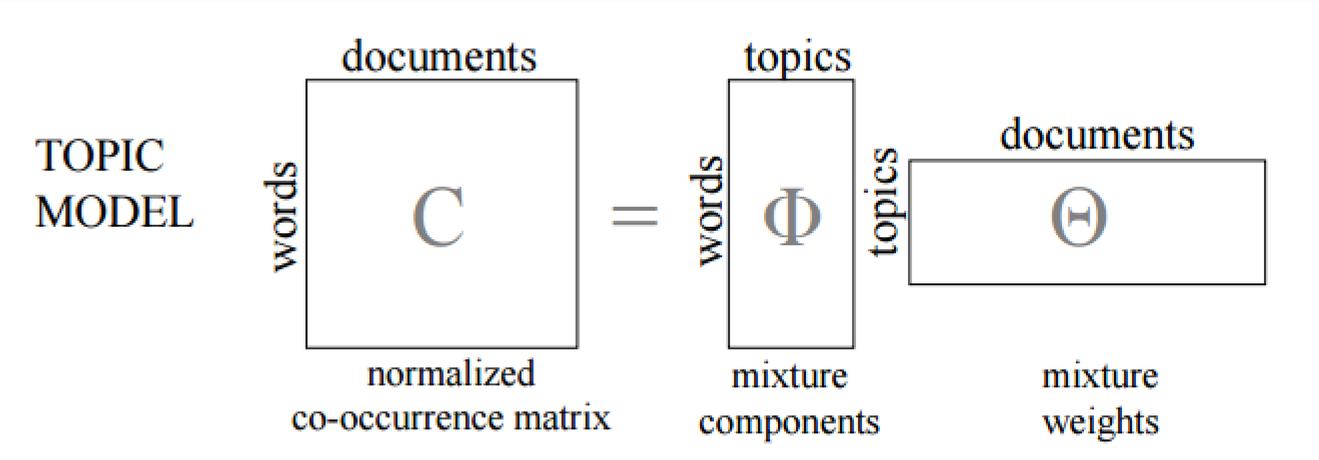
Topic Models

- Hidden topic models model each document in a collection as a distribution over a fixed number of topics
- Each topic characterized by a distribution over a fixed size and discrete vocabulary

- I ate a banana and spinach smoothie for breakfast
- I like to eat broccoli and bananas.
- Chinchillas and kittens are cute.
- My sister adopted a kitten yesterday.
- Look at this cute hamster munching on a piece of broccoli.

- **Sentences 1 and 2**: 100% Topic A
- **Sentences 3 and 4**: 100% Topic B
- **Sentence 5**: 60% Topic A, 40% Topic B
- **Topic A**: 30% broccoli, 15% bananas, 10% breakfast, 10% munching, … (at which point, you could interpret topic A to be about food)
- **Topic B**: 20% chinchillas, 20% kittens, 20% cute, 15% hamster, … (at which point, you could interpret topic B to be about cute animals)

Topic models



Latent Dirichlet Allocation: Notations

- **Notations:**
- M : number of documents
- N: the number of words in a document
- a is the parameter of the Dirichlet prior on the per-document topic distributions,
- β is the parameter of the Dirichlet prior on the per-topic word distribution,
- θ_m is the topic distribution for document m
- φ_k is the word distribution for topic k,
- z_{mn} is the topic for the n-th word in document m, and
- w_{mn} is the specific word.

LDA: Generative Process

- Documents are represented as random mixtures over latent topics, where each topic is characterized by a distribution over words. LDA assumes the following generative process for a corpus D consisting of M documents each of length N_i
- 1. Choose θ_i = Dir(a), where i Є {1,2…M} and Dir(a) is the Dirichlet distribution for parameter a
- 2. Choose φ_k = Dir(β), where k Є {1,2,…K}
- 3. For each of the word positions i,j where j Є {1,2,…N_i }, and i Є{1,…M}
- (a) Choose a topic $z_{i,j}$ = Multinomial(θ_i).
- (b) Choose a word $w_{i,j}$ = Multinomial($\varphi_{zi,j}$)

LDA: Statistical Inference

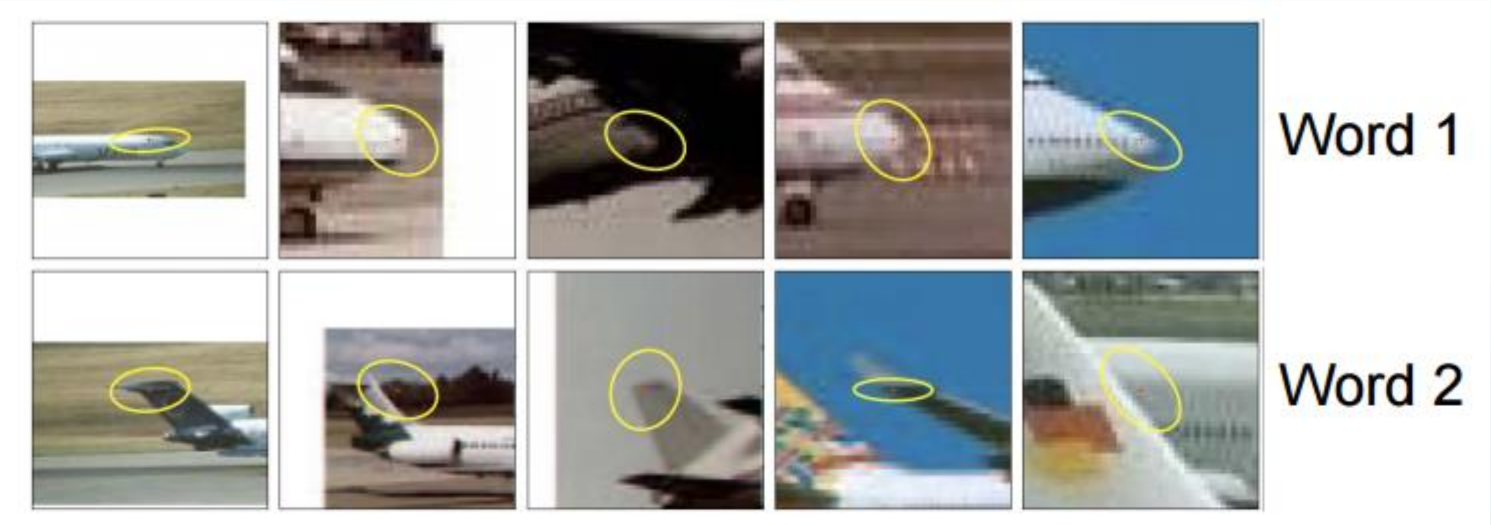
Input: words $\mathbf{w} \in$ documents $\mathbf{d}$
Output: topic assignments $\mathbf{z}$ and counts $n_{d,k}$, $n_{k,w}$, and n_k
begin
    randomly initialize $\mathbf{z}$ and increment counters
    foreach *iteration* **do**
        for $i = 0 \rightarrow N - 1$ **do**
            $word \leftarrow w[i]$
            $topic \leftarrow z[i]$
            $n_{d,topic}$-=1; $n_{word,topic}$-=1; n_{topic}-=1
            for $k = 0 \rightarrow K - 1$ **do**
                $p(z = k|\cdot) = (n_{d,k} + \alpha_k)\frac{n_{k,w}+\beta_w}{n_k+\beta\times W}$
            end
            $topic \leftarrow$ sample from $p(z|\cdot)$
            $z[i] \leftarrow topic$
            $n_{d,topic}$+=1; $n_{word,topic}$+=1; n_{topic}+=1
        end
    end
    return $\mathbf{z}$, $n_{d,k}$, $n_{k,w}$, n_k
end

Algorithm 1: LDA Gibbs Sampling

Extending topic models to image corpus

- Document -> image
- Topic -> object
- Word -> Visual word
- Distribution of hidden topics in an image refers to the degree to which an abstract object such as grass, water, sky, street, etc. is contained in the image.
- Distribution of words in a topic refers to distribution of low-level features in a topic.

Extending topic models to image corpus



Visual bag of words

- Extract feature descriptors(typically SIFT) from each image
- Apply clustering(typically K-means) where K is the dictionary size
- Quantization:

  Cluster centers are the “words” forming a dictionary
- Encoding

  Each SIFT descriptor obtained in the first step is assigned to the “closest” visual word in the vocabulary
- Pooling

  A histogram of visual words of vocabulary is constructed for each document.

Visual bag of words: Querying

- The user input query is also represented as a histogram
- This histogram is compared with those of histograms in the database and the images with least distance using some similarity measures(Eucledian distance, Cosine similarity etc.) are displayed

Topic Models: Retrieving

- Use the document-word matrix obtained from Bag of Words as input to LDA
- Extract z topics from the documents and obtain topic distributions for each document.
- Query image is also represented in this topic space and is compared with topic distributions of other images in the database using some certain similarity measures.
- Closest images are displayed to user

Methodology

- Implementation Language chosen: Python
- Libraries used: Scikit-image

  Numpy

  Matplotlib

  OpenCV

  Scikit-learn
- Subsets of the following 3 different datasets were used for testing:

  Abnormal objects dataset: 623 images belonging to 6 categories

  Caltech 101: 5000 images belonging to 101 categories

  IRMA medical images: 12000 images belonging to 193 categories

Methodology

- SIFT features were extracted using OpenCV's SIFT feature extractor
- Histograms of documents were stored in a file to avoid computing them for every query
- Document-Topic and Topic-word distribution were stored to a file to avoid computing them for every query.
- 2 similarity measures were used and the results of both were compared using: Cosine similarity(CS) and Eucledian distance(EC)

Challenges Faced

- K-Means Clustering: Time and space complexity is high
- Encoding: Assigning each descriptor to a word in the vocabulary is time consuming.
- Time taken by LDA is high

Solutions/ Optimizations

- Used a variation of K-means called mini-batch K-means
- Used OpenCV's interface to FLANN matcher for searching the closest neighbor

FLANN matching

- Fast Library for Approximate Nearest Neighbor Search
- According to the official documentation,

> *"FLANN is a library for performing fast approximate nearest neighbor searches in high dimensional spaces. It contains a collection of algorithms we found to work best for nearest neighbor search and a system for automatically choosing the best algorithm and optimum parameters depending on the dataset."*

Results and discussion

- Parameters:

  Number of retrieved image: 10 and 20

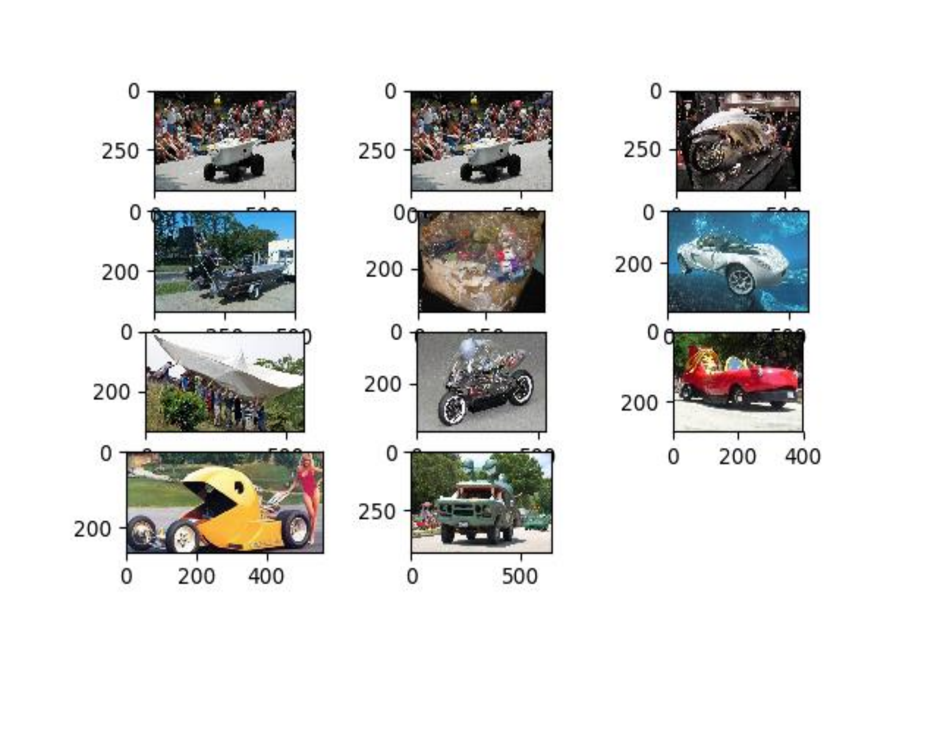
  Number of words in the vocabulary: 2000

  Number of topics: 20

Results and discussion

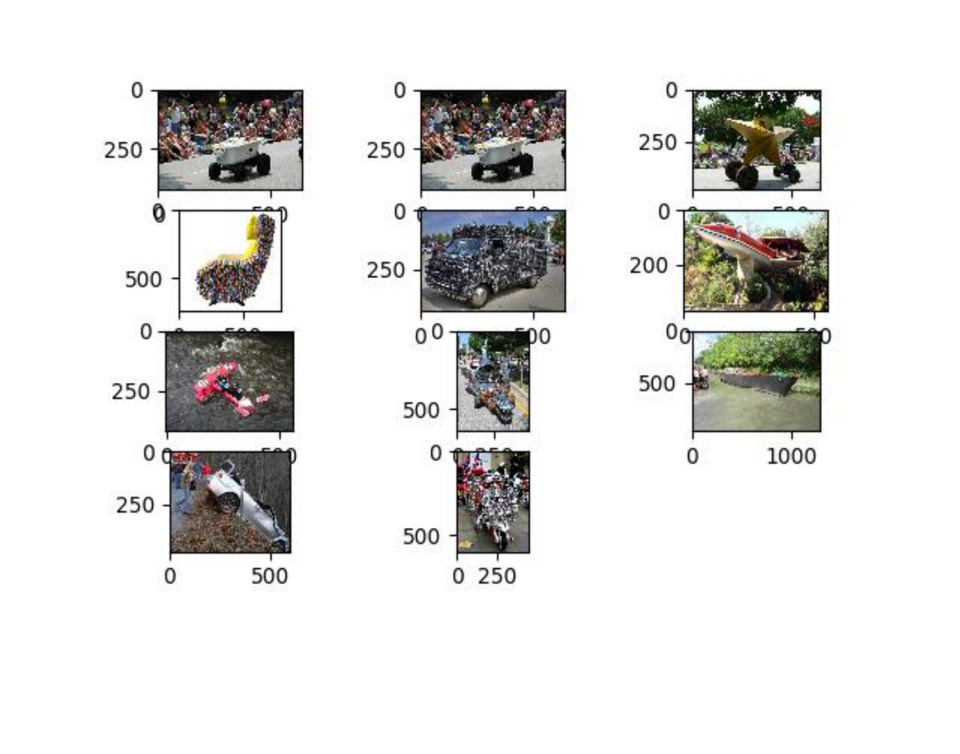
- Abnormal objects: BOW

EC



Precision: 0.4

CS

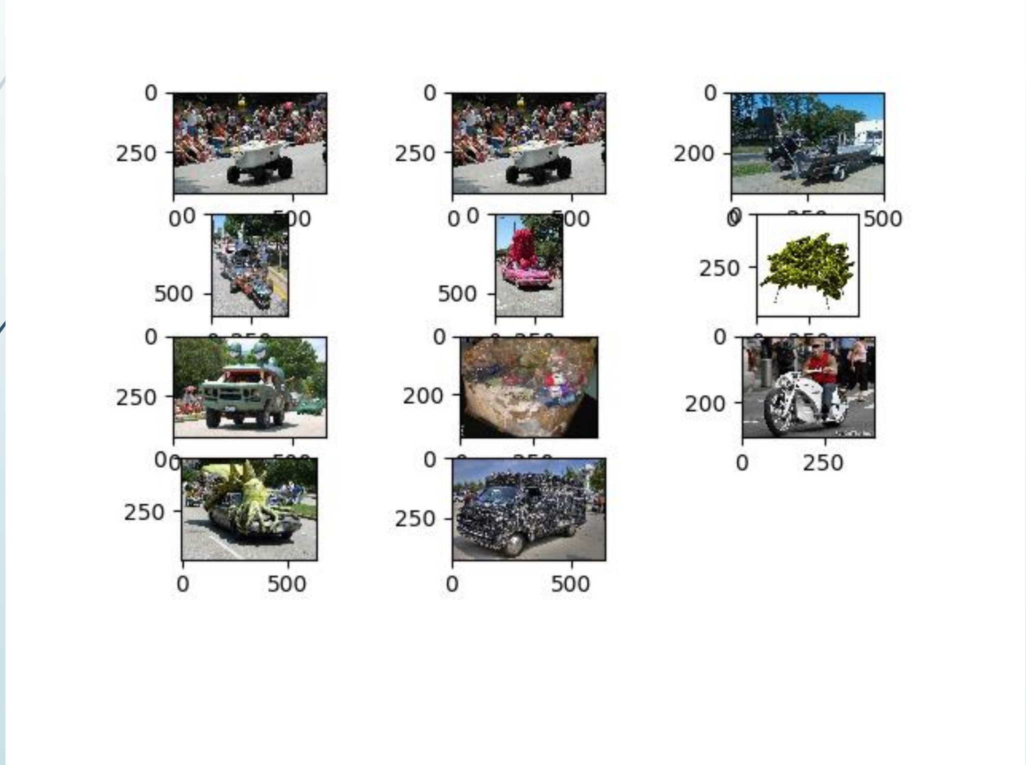


Precision: 0.5

Results and discussion

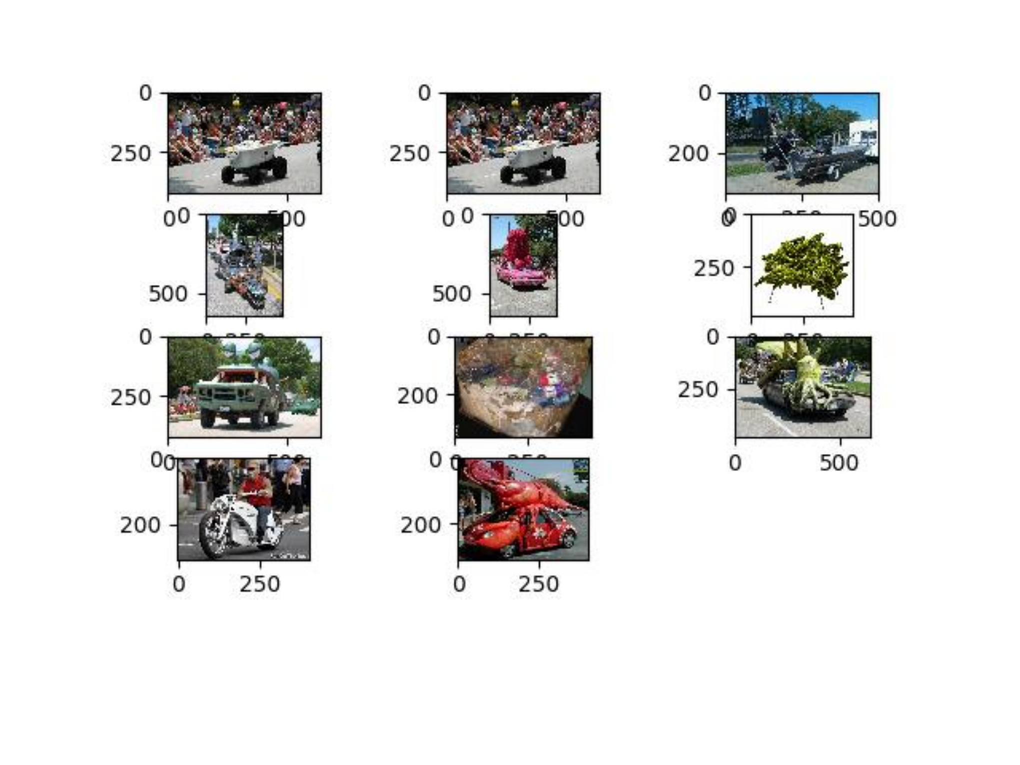
- Abnormal objects: LDA

EC



Precision: 0.6

CS

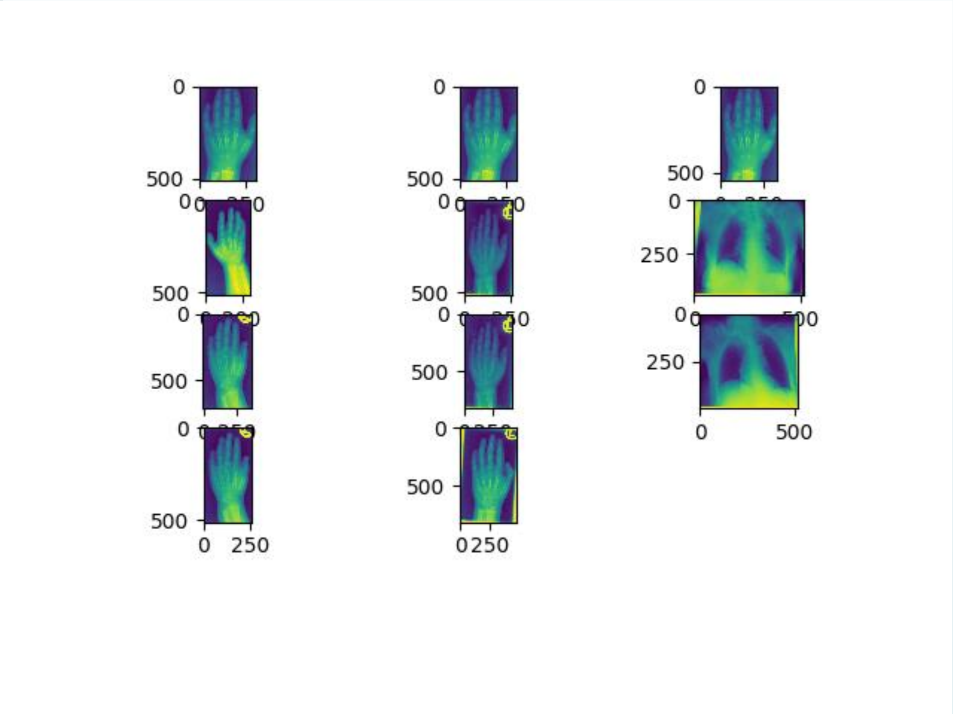


Precision: 0.6

Results and discussion

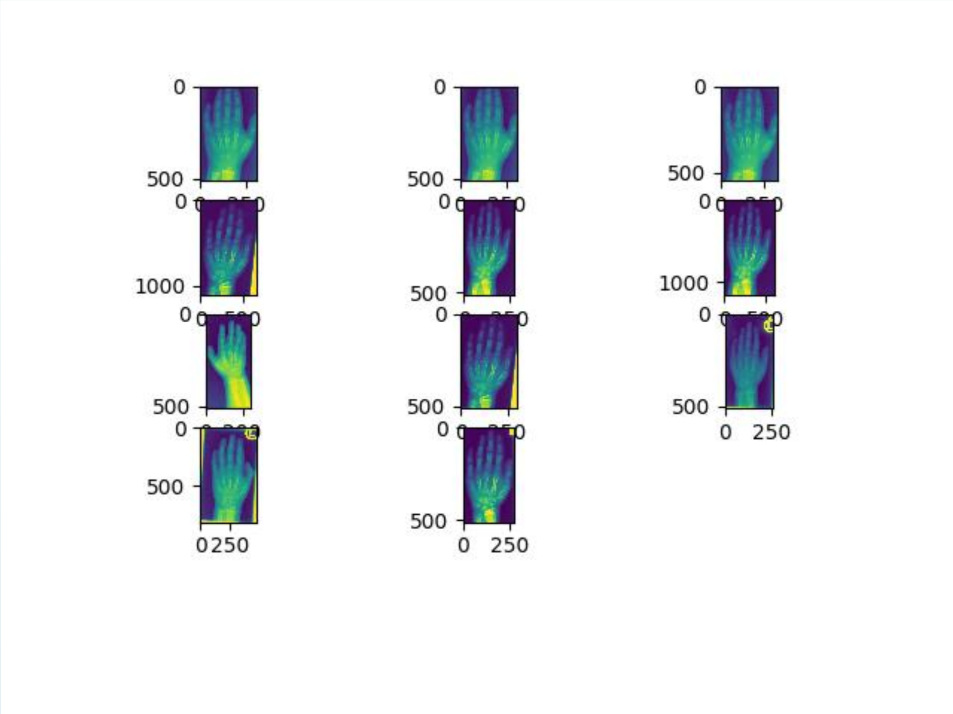
- IRMA dataset: BoW

EC



Precision: 0.8

CS

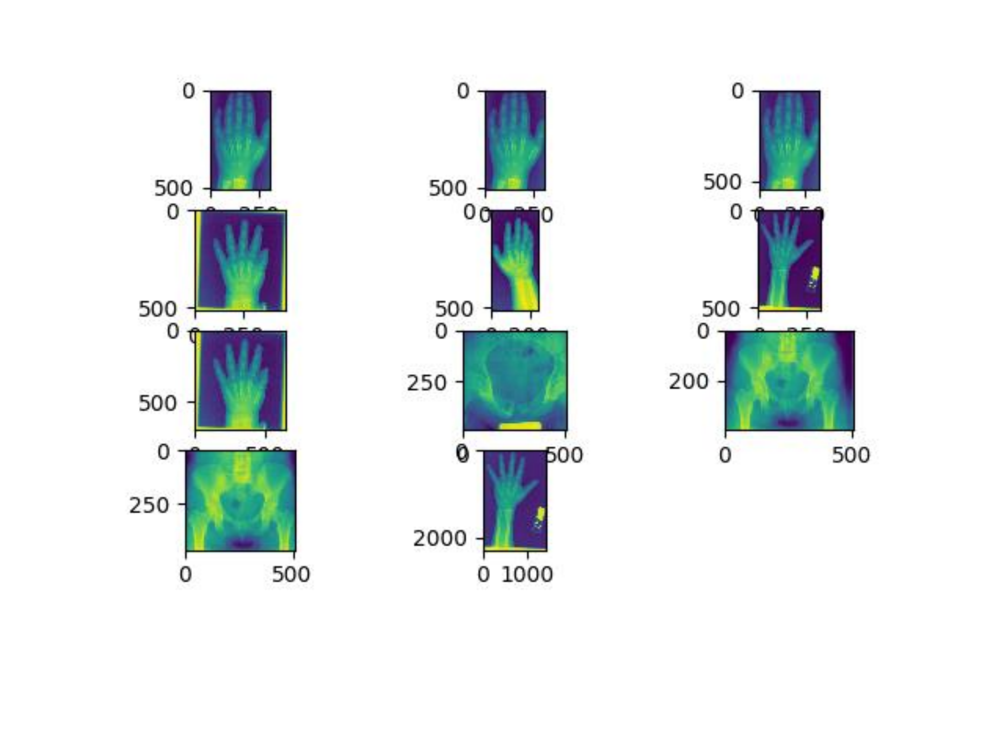


Precision: 1.0

Results and discussion

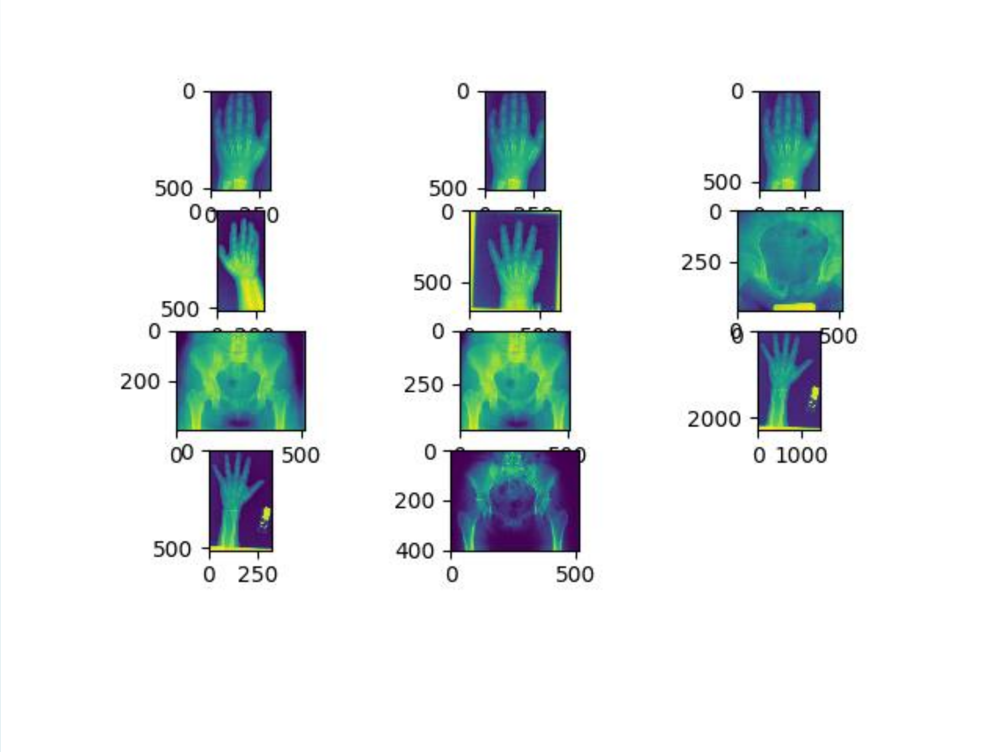
- IRMA dataset: LDA

EC



Precision: 0.7

CS

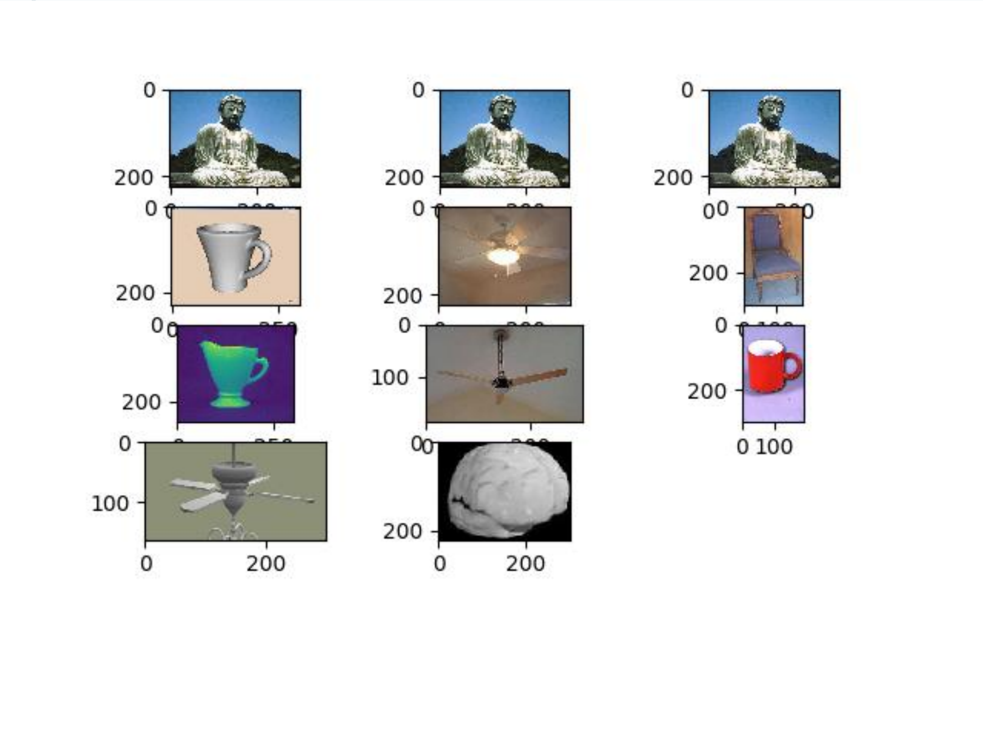


Precision: 0.6

Results and discussion

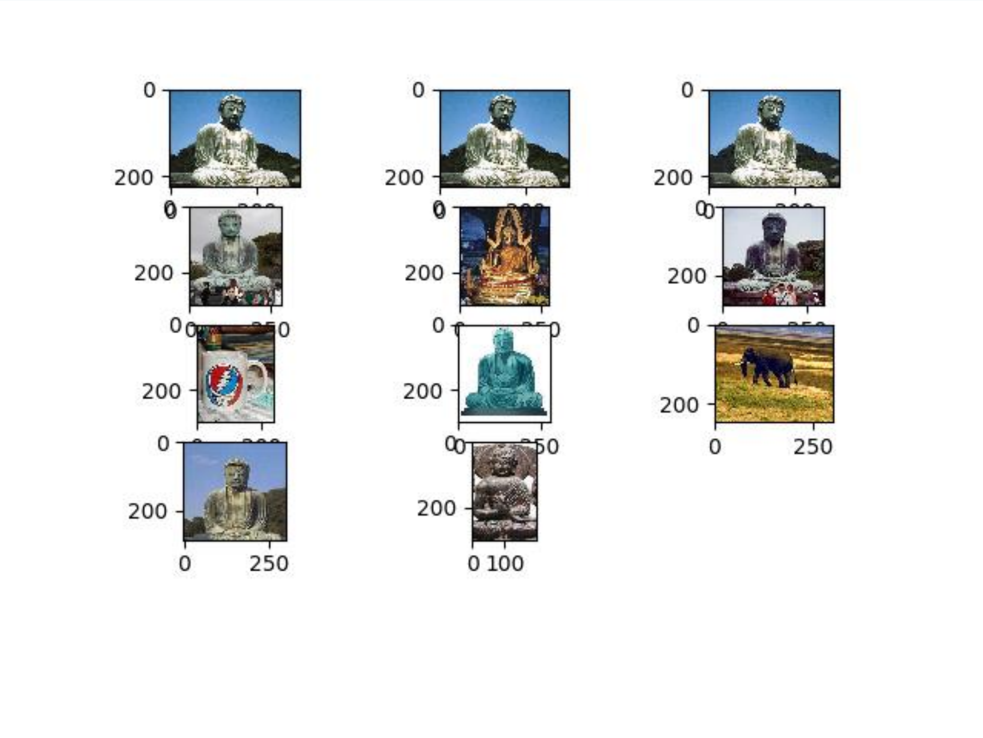
- Caltech 101: BoW

EC



Precision: 0.2

CS

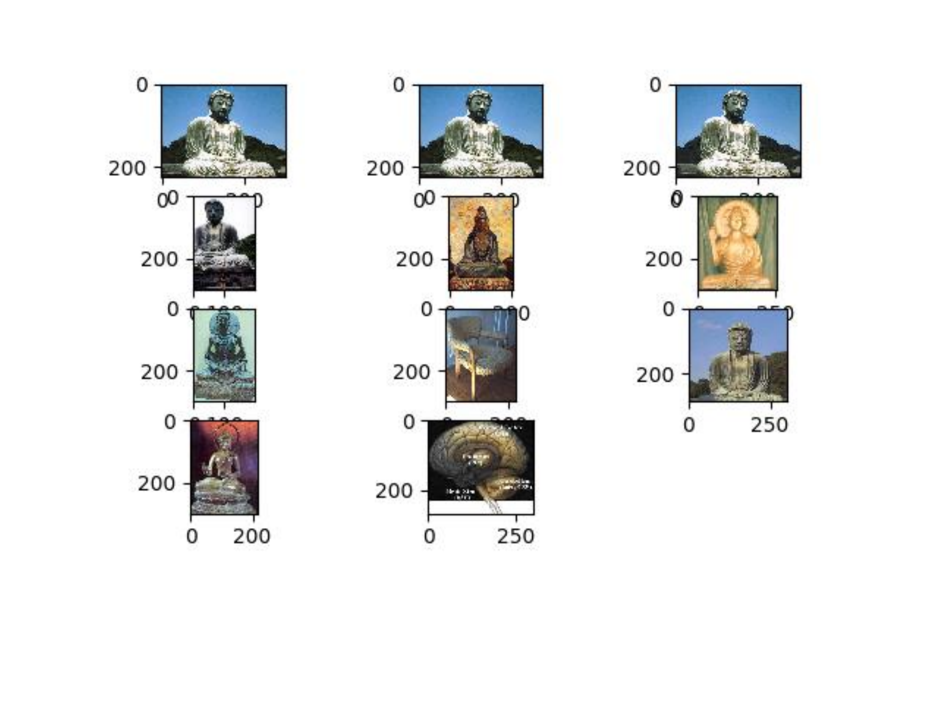


Precision: 0.8

Results and discussion

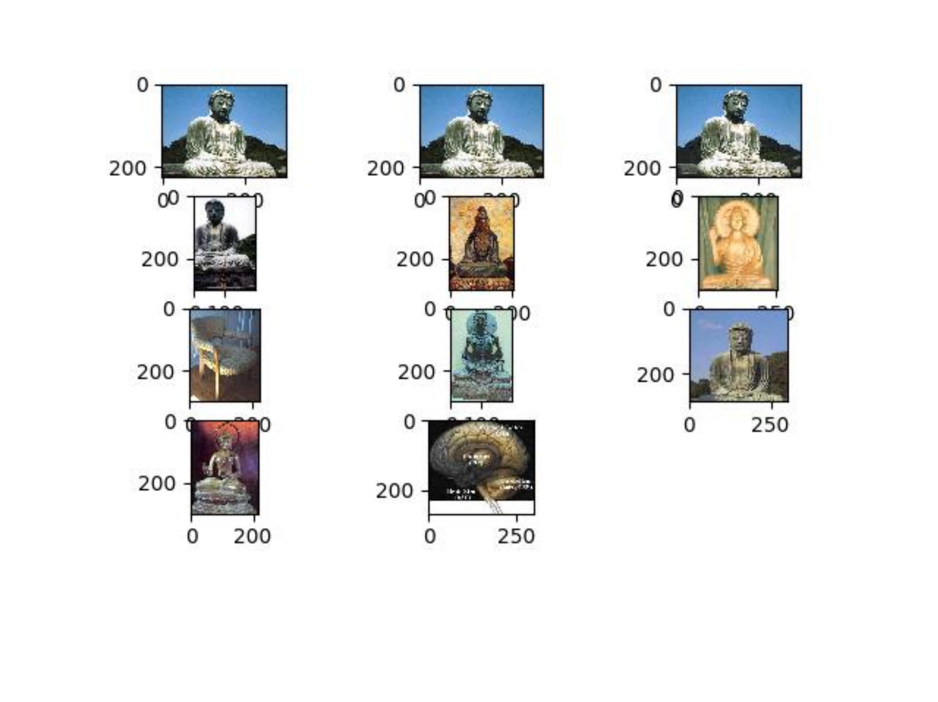
- Caltech 101: LDA

EC



Precision: 0.8

CS



Precision: 0.8

Results and discussion

Summary of results

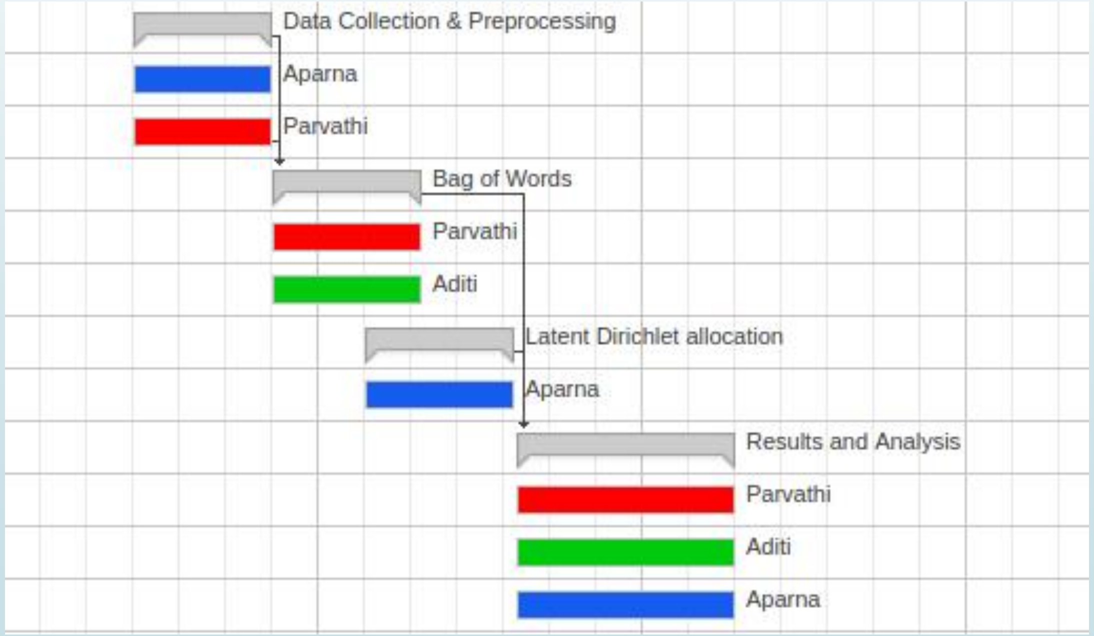
		LDA			BoW	
		EC	CS		EC	CS
	Abnormal	0.5	0.43		0.31	0.38
Precision @ 10	Medical	0.58	0.54		0.4	0.84
	Normal	0.63	0.63		0.16	0.53
	Abnormal	0.34	0.35		0.325	0.29
Precision @ 20	Medical	0.6	0.57		0.36	0.73
	Normal	0.5	0.5		0.12	0.51

Results and discussion

- The average precision @ 10 and precision @ 20 is computed for each dataset with both BoW and LDA and with both EC and CS similarity measures.
- CS clearly performs better than EC in case of BoW.
- In case of LDA, both perform equally well, with EC preforming slightly better.
- In case of Abnormal objects, both precision at 10 and precision at 20 are better for LDA. For medical images, BoW with CS performs much better than LDA. For normal objects dataset, LDA outperforms BoW.

Individual Contribution

- Dataset collection and pre-processing: Aparna PL, Parvathi M.H
- Bag of Words: Aditi Singh, Parvathi M.H
- Latent Dirichlet Allocation: Aparna PL
- Results and analysis: Aparna PL, Aditi Singh, Parvathi M.H



Conclusion and future work

- The topic models use hidden patterns of the local descriptors form the data set and each sample is then represented as the accumulation of the local features of the sample represented in these topics[4].
- LDA was chosen as the topic model and its results were compared with those of BoW retrieval for 3 different datasets. CS was found to outperform EC for BoW. For LDA, EC was found to perform slightly better than EC.
- The future works could be focused on defining a strategy to choose the size of the vocabulary, automatically. For the experiments, SIFT descriptors and LDA model have been used but any other topic model or descriptor could also be used.

References

[1] F.-F. Li and P. Perona. A Bayesian hierarchical model for learning natural scene categories. In *CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2*, pages 524–531, 2005

[2] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman. Learning object categories from google's image search. In *ICCV '05: Proceedings of the Tenth IEEE International Conference on Computer Vision*, pages 1816–1823, 2005

[3] D. M. Blei and M. I. Jordan. Modeling annotated data. In *SIGIR '03: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, pages 127–134, 2003.

[4] Fernandez-Beltran, Ruben, and Filiberto Pla. "Latent topic encoding for content-based retrieval." *Iberian Conference on Pattern Recognition and Image Analysis*. Springer International Publishing, 2015.

Thank you!