Name: Fahad Ahmed Pranto

Id: 162607042

Paper name: **Automated Image Captioning Using Nearest-Neighbors Approach Driven by Top- Object Detections.**

Link: <http://capri.kssk.pwr.edu.pl/CAPRI_2017_Proceedings.pdf#page=6>

**Abstract**: In deep learning, the large efficiency increases combined with the the rapid invention of automatic image captioning systems has contributed to an unprecedented increase of image and video content on the web. In automated image captioning, compositional model-based strategies and retrieval-based methods, two broad concepts have emerged. While in recent times, compositional model-based approaches utilizing the recurrent neural network (RNN) and long short-term memory (LSTM) have seen considerable progress, there are situations in automatic text categorization for which conceptual model-based approaches may not be sufficient and retrieval-based techniques may be more suitable. They introduce an integrated image captioning system that is motivated by top-object identification in order to resolve the technical bottleneck and step up the recovery process.

**Introduction**: Architectures relevant to the area of study of image recognition, such as Artificial Neural Networks (ANN), and Convolution Neural Networks (CNN), in particular, allowed scientists to categorize represents from large collections of results. For a few causes, automated image captioning, i.e., the major issue of presenting in words the situation captured in an image, is understood to be challenging. There are circumstances in the captioning of images that could be best suited to retrieval-based approaches. Examples of cases of this type include: Conditions where there are frequent shifts in the training programs and scenarios in which an automatic image captioning device must be deployed with the objective of which device implementation time and CPU execution time concurrently.

**Literature review**: Even though compositional model-based strategies using the RNN and LSTM are considered to be the state-of-the-art in automatic image captioning, there are possible circumstances for which they might not be well equipped and thus captioning methods based on retrieval may be needed. Retrieval-based approaches to automatic captioning of pictures, therefore, In particular, when an input vector is combined with all the images contained in the database, it can be machine sensitive and slow. Question arises, in what possible conditions will retrieval-based approaches have a benefit over state-of-the-art generative model-based approaches using RNN and LSTM in the form of automatic captioning of images?

**Methods:** To accelerate the caption retrieval process, they used top-object detections to during automatic captioning of images. They use, precisely, the identification of the in order to speed up the k-NN quest for retrieval-based automatic image captioning, the most relevant objects in the image (i.e., the top objects). By detecting all items in a test image, selecting top-n objects (where n is a small number) and extracting all images containing at least one of these n artifacts, in the form of automatic image captioning, results equivalent to those of k-NN retrieval can be obtained by exhaustive study while concurrently obtaining a substantial speedup. Retraining case analysis and complexity analysis are both used in their approach.

**Results:** They use 80 annotated object categories in the MSS for the purpose of training. Dataset COCO. Using VGG-16 fc-7 image attributes, binary SVM classifiers are qualified in each of these 80 annotated classes. When the BLEU and CIDEr similarity measures are used to analyze the obtained captions, the suggested image retrieval, utilizing k-NN search guided by top-object detections, and the traditional image retrieval, which uses exhaustive k-NN scan, produce quite close performance. It is observed that the proposed image retrieval technique based on k-NN search powered by leading-object detection offers an appealing alternative to LSH in order to speed up.

**Conclusion:** They have shown us that they are guided by top-object detections, retrieval-based approaches for automatic image captioning may be made computationally more efficient. In circumstances where the fundamental datasets are evolving continuously, the future benefits of our policy include. Furthermore, the suggested solution requires far fewer parameters Slotting relative to the computer-intensive hyper-parameter tuning associated with conventional k-NN search optimization based on LSH. The method proposed is a suitable candidate for use under solid modelling conditions that often call for Processor time optimization.