IPL at ImageCLEF 2017 Concept Detection Task

Leonidas Valavanis and Spyridon Stathopoulos

Information Processing Laboratory, Department of Informatics, Athens University of Economics and Business, 76 Patission Str, 10434, Athens, Greece valavanisleonidas@gmail.com, spstathop@aueb.com http://ipl.cs.aueb.gr/index_eng.html

Abstract. In this paper we present the methods and techniques performed by the IPL Group for the concept detection task of ImageCLEF 2017. A probabilistic k-nearest neighbor approach was used for automatically detecting multiple concepts in medical images. The visual representation of images was based on the well known, bag of visual words and bag-of-colors models. Detection performance was further enhanced by applying late fusion on the results obtained using different image representations. Our best results were ranked 2nd compared with runs under the same conditions.

Keywords: probabilistic k-nearest neighbors, image annotation, concept detection, quad-tree bag of colors, bag of visual words

1 Introduction

Automatic Image annotation is an important and challenging task within the field of computer vision with applications in several domains. In the medical domain it plays an important role in supporting image search, browsing and organization for clinical diagnosis and treatment. Image retrieval based on semantic information has many advantages and is more robust than using only low-level visual features. In the case of absence of semantic information a typical method to bridge the gap between low level visual features and high level semantics is through the automatic image annotation. This is achieved by applying machine learning techniques to learn a mapping of visual features to textual words. The learned model is then used to assign semantic concepts to new unseen images.

The ImageCLEF caption 2017 task, [3], part of ImageCLEF 2017 [5], consists of 2 subtasks: concept detection and caption prediction. Our group participated in the concept detection subtask. For this task, participating groups were asked to develop systems to identify the presence of relevant bio-medical concepts in medical images.

Details of this task can be found in the overview paper [3] and the web page of the contest ¹. Our approach to concept detection is based on a Probabilistic

¹ http://www.imageclef.org/2017/caption

k-nearest neighbor (PKNN) merging two well known models for image representation, that of the Bag of Visual Words (BoVW), [6] and an improved version of bag of colors (QBoC), [10]. When combined with late fusion, results are further improved, ranking 2nd in best performing runs compared to algorithms that don't rely on external data sources.

The following sections, present the image representation methods and our algorithm for concept detection. Finally we report on our results and conclude with possible venues for further research.

2 Visual representation of images

Three different visual representation models were used in our experiments:

- 1. Localized compact features
- 2. Bag of Visual Words (BoVW)
- 3. Bag of Colors (BoC)

2.1 Localized compact features

Compact visual descriptors have been used extensively in the past years to efficiently represent images in a dataset. For this task, two kinds of visual features were extracted:

- 1. Color and Edge Directivity Descriptor (CEDD)[1].
- 2. Fuzzy Color and Texture Histogram (FCTH)[2].

However, in their original form, these descriptors are extracted globally from an image. In order to include a degree of spatial information, features are extracted over a spatial 4x4 grid. The image is first resized into 256 x 256 pixels and then is split into a 4x4 grid of non-overlaping image blocks. The visual features are then extracted for each block and their corresponding vectors are concatenated to form a single feature vector. The final vector size for the CEDD and FCTH is $4 \times 4 \times 144 = 2,303$ and $4 \times 4 \times 192 = 3,702$ respectively.

2.2 Dense SIFT features and BoVW

Inspired from text retrieval, the Bag-of-visual Words (BoVW) approach has shown promising results in the field of image retrieval and classification. Here the BoVW model was implemented using the DenseSIFT visual descriptor.

The Dense SIFT algorithm [7], is a variant of the SIFT algorithm, which is equivalent to extracting SIFT [8] from a dense grid of locations at a fixed scale and orientation. The SIFT feature is invariant with respect to many common image deformations, including position, scale, illumination, rotation, and affine transformation. The number of features extracted from local interest points using the Dense SIFT descriptor may vary, depending on the image. In order to have a fixed number of feature dimensions, a visual codebook is created by clustering the extracted local interest points of a number of sample images, using the k-means clustering algorithm. After experiments, the number of clusters used is 4,096. Each cluster (visual word), represents a different local pattern which shares similar interest points. The histogram of an image, is created by performing a vector quantization which assigns each key-point to its closest cluster (visual word).

2.3 Quad-Tree Bag-of-Colors Model (QBoC)

The QBoC representation was successfully used for representing images in previous works [9, 10]. With the BoC model [12] a color vocabulary is learned from a sub-set of the image collection. This vocabulary is used to extract the color histograms for each image. The BoC model was used for classification of biomedical images in [4] and it was shown that it is combined successfully with the BoW-SIFT model in a late fusion manner. Similarly to the BoW model the main drawback with the BoC is the lack of spatial information. Quad-Tree decomposition sub-divides an image into regions of homogeneous colors. Each time the image is split into four equal size squares and the process continues until we reach a sub-region of size 1×1 pixel (see Figure 1b). In both models the Term Frequency-Inverse Document Frequency (TF-IDF) weights of visual words were calculated and the image vectors were normalized with the L_1 norm.



Fig. 1. Representation of image 1743-7075-7-33-5 (a) original image; (b) QBoC image.

3 Probabilistic k-NN concept detection (PKNN)

In this section we briefly present our baseline algorithm for automatic concept detection in medical images. The algorithm is divided into two main phases, namely, the visual retrieval step and the annotation step. In the visual retrieval phase, for a given test image, a sample of the k most visually similar images from the training dataset is retrieved. Several experiments on the validation set helped to determine the optimum value for k (k=100).

In the annotation step, the concepts associated to the k retrieved images form the candidate concepts for a test image. The final assigned concepts are determined by a probability score based on the occurrence of concepts in the selected sample. For every distinct concept, w, present in the retrieved training subset, we calculate ConceptScore(w) as:

$$ConceptScore(w) = \sum_{j=1}^{k} P(j) \cdot P(w|j)$$
(1)

where j is an image from the top-k results and P(w|j) is calculated by:

$$P(w|j) = \frac{count(w,j)}{|W_j|} \tag{2}$$

where count(w, j) is the number of times concept w is found in image j, and $|W_j|$ is the total number of concepts in image j. P(j) is considered uniform for all images and thus it is ignored.

The top 6 concepts with the highest *ConceptScore* are selected for the test image. This number was determined by calculating the average number of concepts per image in the training set (i.e. 5.58).

3.1 Concept scoring with Random walks with Restart (RWR)

As an alternative concept scoring method, a Random Walk with Restart (RWR) algorithm [11] was tested. First, from the set of the top k retrieved images an adjacency matrix **A** of size $[c \times c]$ is constructed, where c is the number of distinct concepts in the retrieved images. Matrix **A** defines the graph whose nodes correspond to concepts and edges connect concepts if they are assigned in the same image of the train set. Next, the RWR algorithm is applied to matrix **A** resulting in a vector **r** of size $[c \times 1]$ representing the most probable concepts for the test image. Similarly, the top 6 concepts with the highest r(w) value are selected as the concepts for the test image.

3.2 Late fusion

In order to improve results, late fusion was applied to the ranked lists of concepts for each test image. In late fusion, the ranked concept score lists from different visual descriptors are combined. A new score is calculated based on the combSUM function:

$$combSUM(w) = \sum_{d=1}^{D} ConceptScore_d(w)$$
 (3)

where D is the number of descriptors to combine.

4 Experiments and Submitted Runs

To determine our system's optimal parameters and best visual features we experimented with the validation set provided by the organizers. Trying different visual descriptors and concept detection algorithms, helped us conclude on the submitted runs. Table 1 presents some of the top results obtained from these experiments on the validation set. The "PKNN" prefix in run id corresponds to experiments using the probabilistic concept detection algorithm and "RWR" the random walks with restarts algorithm. The "LFS" prefix corresponds to runs using the late fusion method described in Section 3.2. Table 2 presents the runs submitted to clef and their corresponding results. The same prefixes are also used to describe the individual runs.

Run_ID	Accuracy
PKNN CEDD	0.113
PKNN CEDD 4x4	0.114
PKNN FCTH	0.112
PKNN FCTH 4x4	0.120
PKNN GBOC	0.123
PKNN DSIFT	0.137
RWR GBOC	0.132
RWR DSIFT	0.140
LFS PKNN (FCTH 4x4 DSIFT GBOC)	0.147
LFS PKNN (CEDD 4x4 DSIFT GBOC)	0.145

Table 1. Results on the validation set

Run_ID	Accuracy
LFS_PKNN_DSIFT_GBOC	0.1436
LFS_PKNN_CEDD4x4_DSIFT_GBOC	0.1418
LFS_RWR_DSIFT_GBOC	0.1417
LFS_PKNN_FCTH4x4_DSIFT_GBOC	0.1415
LFS_RWR_CEDD4x4_DSIFT_GBOC	0.1414
DET_LFS_RWR_FCTH4x4_DSIFT_GBOC	0.1394
RWR_DSift_Top100_L2_SqrtNorm_L1Norm	0.1365
PKNN_DSift_Top100_L2_SqrtNorm_L1Norm	0.1364
RWR_GBOC_Top100_L2_SqrtNorm_L1Norm	0.1212
PKNN_GBOC_Top100_L2_SqrtNorm_L1Norm	0.1208

Table 2. Results of the submitted runs with the test set

5 Concluding Remarks

In this report, we presented the image concept detection methods used by the IPL Group for the medical concept detection subtask at ImageCLEF 2017. For

our runs, we used a simple Probabilistic k-Nearest Neighbor approach. Experiments show that using late fusion on BoVW and QBoC performs best and that the image representation plays an important role in performance. Furthermore, the Random walks with Restarts algorithm seemed to perform slightly less, however, a more systematic research is currently underway for this method. Our best run was ranked 2nd in the top performing runs compared to algorithms that don't rely on any external data sources. This results are encouraging and lead to further research on improving the concept detection algorithm with additional textual meta-data.

References

- Chatzichristofis, S.A., Boutalis, Y.S.: Cedd: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. In: ICVS. pp. 312–322 (2008)
- Chatzichristofis, S.A., Boutalis, Y.S.: Fcth: Fuzzy color and texture histogram a low level feature for accurate image retrieval. In: WIAMIS. pp. 191–196 (2008)
- Eickhoff, C., Schwall, I., García Seco de Herrera, A., Müller, H.: Overview of Image-CLEFcaption 2017 - image caption prediction and concept detection for biomedical images. In: CLEF 2017 Labs Working Notes. CEUR Workshop Proceedings, CEUR-WS.org http://ceur-ws.org, Dublin, Ireland (September 11-14 2017)
- García Seco de Herrera, A., Markonis, D., Müller, H.: Bag-of-colors for biomedical document image classification. In: Medical Content-Based Retrieval for Clinical Decision Support, pp. 110–121. Springer (2013)
- Ionescu, B., Müller, H., Villegas, M., Arenas, H., Boato, G., Dang-Nguyen, D.T., Dicente Cid, Y., Eickhoff, C., Garcia Seco de Herrera, A., Gurrin, C., Islam, B., Kovalev, V., Liauchuk, V., Mothe, J., Piras, L., Riegler, M., Schwall, I.: Overview of ImageCLEF 2017: Information extraction from images. In: Experimental IR Meets Multilinguality, Multimodality, and Interaction 8th International Conference of the CLEF Association, CLEF 2017. Lecture Notes in Computer Science, vol. 10456. Springer, Dublin, Ireland (September 11-14 2017)
- Li, F.F., Perona, P.: A bayesian hierarchical model for learning natural scene categories. In: CVPR (2). pp. 524–531 (2005)
- Liu, C., Yuen, J., Torralba, A.: Sift flow: Dense correspondence across scenes and its applications. IEEE Trans. Pattern Anal. Mach. Intell. 33(5), 978–994 (May 2011), http://dx.doi.org/10.1109/TPAMI.2010.147
- Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2. pp. 1150–1157. ICCV '99, IEEE Computer Society, Washington, DC, USA (1999), http://dl.acm.org/citation.cfm?id=850924.851523
- Valavanis, L., Stathopoulos, S., Kalamboukis, T.: IPL at CLEF 2016 medical task. In: Working Notes of CLEF 2016 - Conference and Labs of the Evaluation forum, Évora, Portugal, 5-8 September, 2016. pp. 413-420 (2016), http://ceur-ws.org/ Vol-1609/16090413.pdf
- Valavanis, L., Stathopoulos, S., Kalamboukis, T.: Fusion of bag-of-words models for image classification in the medical domain. In: Advances in Information Retrieval - 39th European Conference on IR Research, ECIR 2017, Aberdeen, UK, April 8-13, 2017, Proceedings. pp. 134–145 (2017), https://doi.org/10.1007/ 978-3-319-56608-5_11

- Wang, C., Jing, F., Zhang, L., Zhang, H.J.: Image annotation refinement using random walk with restarts. In: Proceedings of the 14th ACM International Conference on Multimedia. pp. 647–650. MM '06, ACM, New York, NY, USA (2006), http://doi.acm.org/10.1145/1180639.1180774
- Wengert, C., Douze, M., Jégou, H.: Bag-of-colors for improved image search. In: Proceedings of the 19th International Conference on Multimedia 2011, Scottsdale, AZ, USA, November 28 - December 1, 2011. pp. 1437–1440 (2011), http://doi. acm.org/10.1145/2072298.2072034