

MULTI-PHASE LIVER LESIONS CLASSIFICATION USING RELEVANT VISUAL WORDS BASED ON MUTUAL INFORMATION

Idit Diamant^a Jacob Goldberger^b Eyal Klang^c Michal Amitai^c Hayit Greenspan^a

^aDepartment of Biomedical Engineering, Tel Aviv University, Tel Aviv 69978, Israel

^bFaculty of Engineering, Bar-Ilan University, Ramat-Gan, Israel

^cDepartment of Diagnostic Imaging, The Chaim Sheba Medical Center, Tel-Hashomer, Israel

ABSTRACT

We present a novel method for automated diagnosis of liver lesions in multi-phase CT images. Our approach is a variant of the Bag-of-Visual-Words (BoVW) method. It improves the BoVW model by selecting the most relevant words to be used for the input representation using a mutual information based criterion. Additionally, we generate relevance maps to visualize and localize the decision of the automatic classification algorithm. We validated our algorithm on 85 multi-phase CT images of 4 categories: hemangiomas, Focal Nodular Hyperplasia (FNH), Hepatic Cellular Carcinoma (HCC) and cholangiocarcinoma. The new algorithm suggested in this paper improves the classical BoVW method sensitivity by 16% and specificity by 3%. The shift from single-phase liver data to a multi-phase representation is shown to substantially improve classification results. Overall, the system presented reaches state-of-the-art classification results of 80% sensitivity and 92% specificity on the 4 category lesion data, a challenging clinical diagnosis task.

Index Terms— Liver lesions, automated diagnosis, visual words, mutual information, classification, feature selection.

1. INTRODUCTION

Cancer is a leading cause of death worldwide. Among the predominant cancer types, liver cancer ranks at fourth place, accounting for more than 600,000 deaths each year. The number of liver tumors diagnosed throughout the world is increasing at an alarming rate. Early diagnosis and treatment is the most useful way to reduce cancer deaths.

Multi-phase contrast-enhanced Computed Tomography (CT) is the primary imaging technique employed by clinicians for detection, diagnosis, and monitoring of liver lesions. Multi-phase data includes: unenhanced (non-contrast) scan, arterial phase scan (30-40 seconds after contrast injection), portal venous phase scan (70-80 seconds after injection) and delayed phase scan (3-5 minutes after injection). Liver lesions show different visual characteristics at the different scans. Examples of lesions at the different scans are shown

in Fig. 1. These visual features carry important diagnostic information which is used to classify the lesions into benign or malignant categories.

Liver lesions differentiation is a difficult task due to large variability in lesions appearance. Moreover, the imaging appearance of different lesions overlaps, leading to considerable inter-reader variations. Diagnoses is based on physicians subjective evaluation and is dependent on the physicians experience. Although in recent years the quality of CT images has been significantly improved, often radiologists confirm their diagnosis by using tissue biopsies. Thus, there is a need for automated diagnostic tools to assist radiologists in evaluating liver lesions.

A number of researchers have developed methods for automated diagnosis of liver lesions. Works on CT liver lesion classification have mainly focused on single-phase data. Recently works have started to emerge using multi-phase data [1, 2, 3, 4]. Quatrehomme et al. [1], for example, performed classification of cysts, adenomas, haemangiomas, HCC and metastasis and concluded that there is a significant improvement using multi-phase images compared to single-phase, obtaining ~75% average precision and average recall using the multi-phase data. One, well-known machine learning model used by the machine vision community is the Bag-of-Visual-Words (BoVW) model [2]. The BoVW method and its variants have been shown to be successful in medical classification tasks (e.g. [2, 5]).

In this work, we present a novel variant of the BoVW model for automatic classification of liver lesions in multi-phase CT images. In the classical BoVW approach, the image is divided into image patches, from which visual words are computed. The visual words from all images in the collection are assembled into a dictionary, from which feature vectors are created for each image (based on a histogram of visual word occurrences in the image). In the current work we propose to select a subset of the words in the dictionary, as the most relevant to the task, and we suggest that using such a subset can improve performance. Our algorithm, BoVW-MI, is based on selection of most relevant words from a generated dictionary, based on a mutual information (MI) criterion. The

same approach is also used for localized lesion visualization. We present the BoVW-MI model in Section 2. Experimental results using the model are shown in Section 3.

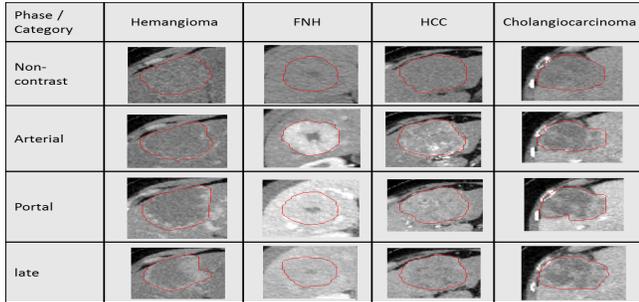


Fig. 1. Examples of multi-phase liver lesions data. Radiologist’s lesion markings are shown in red.

2. FEATURE RELEVANCE BASED ON A MUTUAL INFORMATION CRITERION

2.1. BoVW model

The first step of dictionary generation in the classical BoVW approach is a patch extraction phase, in which we extract uniform size patches from each region-of-interest (ROI). Once patches are defined, each patch is represented by a set of feature descriptors. To reduce the computational complexity of the algorithm and the level of noise, a principal component analysis procedure (PCA) is applied to the patch collection. The final step includes a k -means clustering step on the vectors of the patch collection, to generate a representative visual-word dictionary. A given (training or testing) image x can now be represented by a unique distribution over the generated dictionary of words, which is a representative image histogram (x_1, \dots, x_k) . x_i is the relative number of image patches in x that were mapped to the i -th visual word based on content similarity. These k numbers are the features extracted from the image to be used in the classification algorithm, where each feature corresponds to a visual word. Following the representation phase, various classifiers can be used for classification.

2.2. Measuring the Relevance of words to a specific task

In the current work, once a visual dictionary is learned, we find the *relevance* of each feature (word) to the specific classification task at hand. Feature *relevance* is often characterized in terms of mutual information between the feature value and the class label. We compute the mutual information between the image label and each of the features as follows: Suppose we are given n images with labels c_1, \dots, c_n and the feature vector of the t -th image is denoted by (x_{t1}, \dots, x_{tk}) . To obtain a reliable estimation of the mutual information for each

feature i , we first quantize the i -th feature values across all the images x_{1i}, \dots, x_{ni} into L levels (in our implementation we sort the n values and divide them into five groups of equal size). Denote the quantized version of x_{ti} by $y_{ti} \in \{1, \dots, L\}$. Denote the joint probability of the (quantized) i -th feature and the image class by:

$$p_i(v, c) = \frac{1}{n} |\{t | y_{ti} = v, c_t = c\}| \quad (1)$$

where c is the class label, v is the quantized bin level and $|\cdot|$ is the set cardinality. The mutual information between the class label variable C and the quantized feature Y_i is:

$$I(Y_i; C) = \sum_{v=1}^L \sum_c p_i(v, c) \log \frac{p_i(v, c)}{p_i(v)p(c)} \quad (2)$$

where $p_i(v)$ and $p(c)$ are the marginal distributions of the i -th feature and the class label respectively. In our case C is a four-valued random variable and therefore, $0 \leq I(Y_i; C) \leq 2$.

The mutual information computed using Eq. 2 is the *relevance* of each feature (each visual word). Using MI values of each word, we can now generate *relevance maps* for each ROI. These maps are created by placing in each image patch center the mutual information of the visual word it is assigned to. These maps demonstrate how the visual word methodology works and show the spatial layout of the most relevant words. This helps us to visualize and localize the area within the ROI that is most relevant to the diagnostic task.

2.3. Relevant words Selection

It can be hypothesized that only a selection of the words in the dictionary is in fact relevant to a specific task at hand (and differs per task). A question arises whether a selection of a subset of words, the ones most relevant to the task, can increase performance. In our BoVW-MI model, we define a set of selected visual words as those that have high mutual information values (larger than a preselected threshold parameter). These are the most relevant words for the classification task. The input image histogram is now reduced to contain only selected words, resulting in a reduced dimension feature vector. This new representation is next used for the classification task.

3. EXPERIMENTAL RESULTS

We evaluated our algorithm on a dataset of multi-phase CT images of 85 liver lesions: 27 hemangiomas (HEM), 16 FNH, 29 HCC and 13 cholangiocarcinoma (CLG), acquired from the Sheba medical center. All 4 phases were collected for each lesion. Cases were acquired between 2011 and 2013 using GE Medical Systems scanner. Cases were collected with approval of the institutions Institutional Review Board. Fig. 1 shows examples of marked lesions of each type and phase.

Table 1. Performance comparison between BoVW-MI and BoVW (four phase input).

	BoVW-MI		BoVW	
	Sensitivity [%]	Specificity [%]	Sensitivity [%]	Specificity [%]
HEM	88.9	89.7	63.0	91.4
HCC	86.2	91.1	65.5	91.1
FNH	87.5	97.1	75.0	95.7
CLG	38.5	94.4	46.2	75.0
weighted average	80.0	92.3	63.5	89.6

The intensity of liver lesions in CT images is important for diagnosis. Thus, for the task at hand we use the raw image pixels as patch descriptor. We use the support vector machine (SVM) classifier with histogram intersection kernel on normalized feature vectors. Histogram intersection has no free kernel parameters, which makes it convenient for fast parameter evaluation. For multi-phase data input, we generated a different dictionary for each phase and represented each lesion with a concatenated histogram of all phases. Evaluation was performed using a leave-one-out cross-validation method.

3.1. Visualization of relevant areas

In multiphase liver lesion classification, radiologists search for visual characteristics with specific patterns along the phases to differentiate between the categories. Hemangiomas, for example, appear hypodense in the arterial phase in the lesion interior region with peripheral nodular enhancement, in contrast to HCC and FNH. Cholangiocarcinomas are also hypodense in the lesion center, however they are less homogenic than hemangiomas. FNH demonstrates bright arterial contrast enhancement except for the central scar which remains hypodense. In the portal venous phase the lesion becomes isodense to liver.

We extracted liver lesion relevance maps, as described in Section 2.2. Example results of relevance maps are shown in Fig. 2. By observing these maps we can learn how the algorithm works. For example, in the arterial phase, the interior region is most relevant for differentiation between the categories and the boundaries are less relevant. Additionally, FNH arterial maps show relevant areas (bright) located at the lesion interior, except for the scar area which have similar intensity to parenchyma tissue. On the other hand, in the portal phase high MI values are assigned only to scar area, which is the area that characterizes FNH in this phase.

3.2. Classification performance evaluation

The performance of the proposed method was evaluated using single and multi-phase images. We tested the algorithm performance for different parameter values: patch sizes varying between 5 and 13 pixels (width), visual word size (amount of PCA coefficients) between 10 and 16, visual word amount

in dictionary (dictionary size) between 80 and 800 for each phase (see Fig 3a) and MI threshold between 5 to 99.5% of words removed (leaving only high MI words). We optimized to obtain a combination of parameters which achieves the best performance. The algorithm performance was evaluated using sensitivity and specificity measures.

Optimal parameters of classical BoVW were found to be patch size of 11, word size of 10 PCA coefficients and dictionary size of 100 words per phase (400 words total). Using these parameters, optimal MI threshold for BoVW-MI was found to be 90% (% of words to be removed), leaving 40 words. Optimal weighted average sensitivity and specificity values of 80% and 92.3% respectively achieved using BoVW-MI using all 4 phases. Using classical BoVW reached optimal sensitivity of 63.5% and specificity of 89.6%. Table 1 shows sensitivity and specificity results for each category. Fig. 3a shows a comparison of classification accuracy (weighted average sensitivity) levels between BoVW-MI using the subset of words with high MI values, BoVW-MI using the complement set of words with low MI values and BoVW (that is based on all the dictionary words), for different dictionary sizes. Fig. 3b shows a comparison between usage of the subset of words with high MI values and the complement subset of words with low MI values, for different threshold levels, using optimal parameters combination. These figures show that using relevant words (with high MI) improves performance compared to classical BoVW for each dictionary size. On the other hand, usage of remaining words in dictionary (with low MI) results in lower accuracy levels. By using relevant words representation, we remove noise introduced by non relevant areas, thus improving performance. However, when using non relevant (low MI) representation, we are left with mainly noise which lead to low classification accuracy.

Additionally, we investigated the influence of multi versus single phase data. Table 2 shows performance comparison using BoVW-MI between single and multi phase data. We used patch size of 11 and word size of 10, as found to be optimal as shown in the previous section. We optimized the dictionary size parameter for each input separately (single/multi data), obtaining optimal dictionary of 4×100 for multi-phase data and 280 for single-phase data. Multi-phase representation improves the weighted average sensitivity by 10.6% and specificity by 3.6% compared to single phase representation.

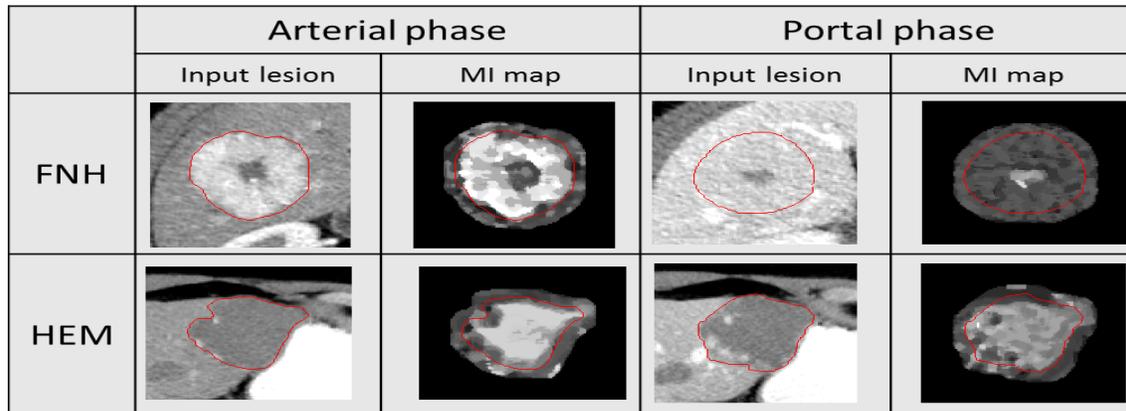


Fig. 2. Examples of relevance maps of liver lesions. High intensity (bright areas) correspond to high MI values, meaning areas which are more relevant to the task. Maps MI values for ROI only (non used background is shown in black). Radiologist’s lesion markings are shown in red.

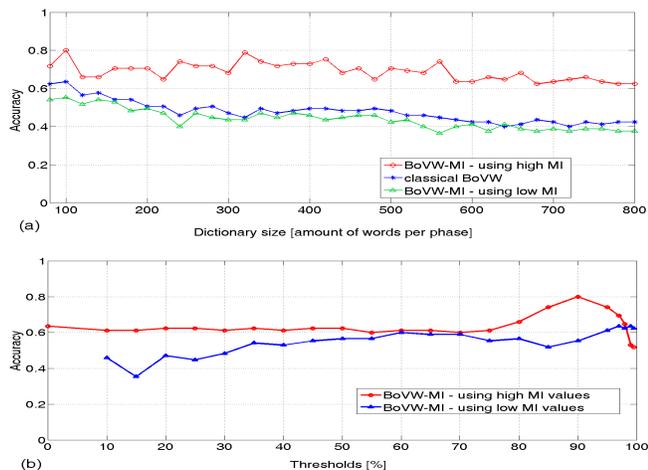


Fig. 3. Comparison between BoVW-MI using words with high MI values and using words with low MI values. (a) Comparing to classical BoVW for different dictionary sizes; (b) Using optimal parameter combination. When using high MI - the threshold refers to % of words removed. For low MI - it refers to % of words left.

Table 2. Single vs. multi-phase using BoVW-MI.

Phases	Average Sensitivity	Average Specificity
Portal	69.4%	88.7%
All	80.0%	92.3%

To conclude, in this study we presented a novel approach for liver lesion diagnosis of hemangioma, HCC, FNH and cholangiocarcinoma in multi-phase CT images. Our method selects visual words that are most relevant for the classification task such that the relevance is measured as the MI be-

tween the visual word and the image class. We showed that the shift from single-phase liver data to a multi-phase representation improves classification performance. In addition, we demonstrated that classification based on an informative selected set of words, per task, results in a significant performance improvement. This is a general method that can be used in additional medical classification tasks.

4. REFERENCES

- [1] A. Quatrehomme, I. Millet, D. Hoa, G. Subsol, and W. Puech, “Assessing the classification of liver focal lesions by using multi-phase computer tomography scans,” *MCCR-CDS*, pp. 80–91, 2012.
- [2] W. Yang, Z. Lu, M. Yu, M. Huang, Q. Feng, and W. Chen, “Content-based retrieval of focal liver lesions using bag-of-visual-words representations of single- and multiphase contrast-enhanced ct images,” *J. Digit. Imaging*, vol. 25, pp. 708–719, 2012.
- [3] D. Duda, M. Krtowski, and J. Bzy-Wendling, “Computer-aided diagnosis of liver tumors based on multi-image texture analysis of contrast-enhanced CT, selection of the most appropriate texture features,” *Studies in Logic, Grammar and Rhetoric*, vol. 35, pp. 49–70, 2013.
- [4] S. Roy, Y. Chi, J. Liu, S. Venkatesh, and M. Brown, “Three-dimensional spatio-temporal features for fast content-based retrieval of focal liver lesions,” *IEEE Transactions on Bio-Medical Engineering*, vol. 92, pp. 1–10, June 2014.
- [5] I. Diamant, J. Goldberger, and H. Greenspan, “Visual words based approach for tissue classification in mammograms,” in *Proceeding of SPIE Medical Imaging*, vol. 8670, 2013.