Semantic Description of Medical Image Findings: Structured Learning Approach

Pavel Kisilev[1]
pavelki@il.ibm.com

Eugene Walach[1]
walach@il.ibm.com

Sharbell Hashoul[1,2]
sharbelh@il.ibm.com

Ella Barkan[1]
ella@il.ibm.com

Boaz Ophir[1]
boazo@il.ibm.com

Sharon Alpert[1]
asharon@il.ibm.com

[1] IBM Haifa Research Lab, Haifa, Israel

[2] Carmel Medical Center, Haifa, Israel

Computer Aided Diagnosis (CADx) systems are designed to assist doctors in medical image interpretation. However, a CADx is often thought of as a "black box" whose diagnostic decision is not intelligible to a radiologist. Therefore, a system that uses semantic image interpretation, and mimics human image analysis, has clear benefits. In this paper, we propose a system which automatically generates textual description of medical image findings, such as lesions.

Having found a lesion, a radiologist examines its visual appearance characteristics to make a final diagnosis. The visual appearance is usually described in terms of semantic descriptors such as: shape, orientation, margin, boundary type, contrast enhancement, localization, mass effect on surrounding tissues, and others.

The estimation of semantic descriptor values requires explicit or implicit representation by a diverse set of image measurements that describe each one of the semantic descriptors quantitatively. We use various image measurements to calculate the informative features, such as histograms of pixel values, shape and texture descriptors and others.

We pose the problem of semantic description of a lesion as learning to map a set of image based informative features to a set of semantic descriptor values. A lesion is described by a set of J semantic descriptors. Semantic description of the i-th lesion is an assignment $y_i = [y_{i,1}, \ldots, y_{i,J}]$ where each j-th semantic descriptor $y_{.,j}$ can have one of the possible discrete values $Y_j \in \{1, \ldots, V_j\}$ corresponding to the radiological lexicon. Following the standard practice in structured learning, the energy function of the above assignment is a sum of unary and pair-wise terms [1]:

$$E(\mathbf{y}_i) = \sum_j \mathbf{u}_1^T \phi_1(y_{ij}, \mathbf{X}_i) + \sum_{j,k \in \mathcal{S}} \mathbf{u}_2^T \phi_2(y_{ij}, y_{ik}, \mathbf{X}_i), \quad (1)$$

where $\mathbf{X}_i$ are image measurements (visual features); ϕ_1 are the unary potentials that capture the relationship between the image measurements and the semantic descriptor values; ϕ_2 are the pairwise potentials that capture joint relationships between the semantic descriptors, and reflect the likelihood of semantic descriptors to jointly have particular values; S is the set of all possible pairs of semantic descriptors; $\mathbf{u_1}, \mathbf{u_2}$ are the model parameters.

We learn the parameters $\mathbf{u} = [\mathbf{u}_1^T \ \mathbf{u}_2^T]^T$ of the model (1), using Structured SVM (SSVM) framework. In particular, given N training examples, the model parameters are learned by optimizing the regularized large-margin objective [1], [2]:

$$\hat{\mathbf{u}} = \min_{\mathbf{u}, \xi \geq 0} \frac{1}{2} \|\mathbf{u}\|^2 + C\xi$$
$$\text{s.t.} \quad \frac{1}{N} \sum_{i=1}^{N} \max_{\bar{\mathbf{y}}_i \in \mathcal{Y}} \left[\Delta(\bar{\mathbf{y}}_i, \mathbf{y}_i^*) - \langle \mathbf{u}, \psi(I_i, \bar{\mathbf{y}}_i) \rangle + \langle \mathbf{u}, \psi(I_i, \mathbf{y}_i) \rangle \right] \leq \xi \quad (2)$$

where the unary and the pairwise potentials are concatenated into a column vector ψ. Once the model parameters are learned, the inference goal is, given a new lesion, to find the best assignment whose semantic values result in the lowest energy (1). This is achieved by:

$$\hat{\mathbf{y}}_i = \arg\min_{\bar{\mathbf{y}}_i \in \mathcal{Y}} < \hat{\mathbf{u}}, \psi(\bar{\mathbf{y}}_i, \mathbf{X}_i) >, \quad (3)$$

whose solution is found using the Sequence Alignment algorithm [2].

The main mode of operation of the proposed method is illustrated in

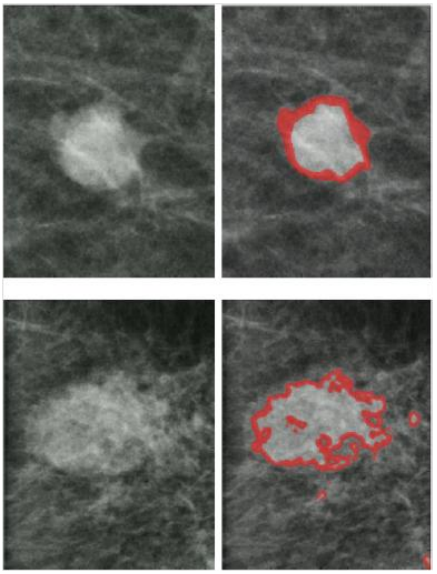

Figure 1: Examples of automatically generated textual description of lesions, using the proposed method: (a) region of interest (ROI) taken from a breast mammography image, (b) found lesion contour (c) textual description of the lesion (estimated semantic descriptors are marked in yellow colour).

Figure 1. Two examples of input images, of the detected lesion contours, and of the corresponding automatically generated textual descriptions of lesions using the proposed method are depicted in Figure 1a-1c, respectively. Given a medical image, the first step is to localize a lesion and to find its contour. In this work, we concentrate on the problem of automatic generation of semantic description of lesions. We, therefore, use semi-automatic lesion detection and contour extraction, instead of a fully automatic approach. We assume that the bounding box around a lesion is found or given by a radiologist. Then, an active contour type method is applied to find the lesion contour inside of the bounding box. Given the found lesion contour, we calculate various image measurements and, based on it, construct visual features. Finally, we use the learned in advance model of mapping from visual features to semantic values, and generate the semantic description of a new lesion.

The proposed approach generates radiological lexicon descriptors used to make a diagnosis of various diseases. This can help radiologists easily understand a diagnosis recommendation made by an automatic system, such as CADx. We apply the proposed method to publicly available and to proprietary breast and brain imaging datasets, and show that our method generates more accurate descriptions, as compared to other alternative approaches.

[1] Nowozin, Sebastian, and Christoph H. Lampert. "Structured learningand prediction in computer vision." Foundations and Trends in Computer Graphics and Vision 6.3–4 (2011): 185-365.

[2] I. Tsochantaridis, T. Joachims, T. Hofmann, and Y. Altun. Large Margin Methods for Structured and Interdependent Output Variables. Journal of Machine Learning Research (JMLR), 6(Sep):1453-1484, 2005.