Mapping Visual Features to Semantic Profiles for Retrieval in Medical Imaging

Johannes Hofmanninger, Georg Langs
Department of Biomedical Imaging and Image-guided Therapy
Computational Imaging Research Lab, Medical University of Vienna
www.cir.meduniwien.ac.at

Content based image retrieval (CBIR) is highly relevant in medical imaging, since it makes vast amounts of imaging data accessible for comparison during diagnosis. Finding image similarity measures that reflect diagnostically relevant relationships is challenging, since the overall appearance variability is high compared to often subtle signatures of diseases. To learn models that capture the relationship between semantic clinical information and image elements at scale, we have to rely on data generated during clinical routine (images and radiology reports), since expert annotation is prohibitively costly.

The problem of learning relations between local image regions and labels on the image level can be posed as a multi-label multi-instance learning (MIL) problem. Retrieval related to clinical findings such as lung textures poses a very particular form of MIL different to standard MIL metric learning techniques in several aspects. The number of instances in the bags is substantially higher compared to standard MIL data reported in literature ($\gg 1000$ vs. ~ 10 as in e.g. [1, 2]) or MI benchmark datasets such as *Fox, Tiger, Elephant*. The optimization problem in [1] grows quadratically with the number of instances. Furthermore, when analysing medical imaging data, the bags are heavily skewed, each bag containing a large portion of healthy instances since even patient lungs contain healthy tissue. This poses challenges to distance definitions on the bag level where the minimum distance among the instances of two bags is used to judge their relationship [1, 2].

We demonstrate that re-mapping visual features extracted from medical imaging data based on weak image volume level label information creates descriptions of local image content that capture clinically relevant information. These labels can be extracted from radiology reports that describe findings in image volumes. Results show that these features enable higher recall and precision during retrieval compared to visual features. Furthermore, after learning, we can map specific semantic terms describing disease patterns to localized image volume areas.

The method consists of a training and an indexing- or application phase. During training multiple dense, random, independent partitionings of the feature space are generated by a random ferns ensemble [3]. Based on the label distributions in the resulting partitions, a remapping of feature vectors is generated that captures the link between appearance and weak labels. In the indexing- or application phase, an ensemble affinity for a novel record to each class is calculated, and a corresponding semantic profile feature vector is generated.

For the experimental validation we created a weakly labeled data set from the labeled lung data. We performed experiments on a set of 300 high resolution computed tomography (HRCT) scans of lungs. All voxels in the images are labelled into one of five tissue classes: healthy lung texture and four types (ground-glass, reticular pattern, honey combing, emphysema) occuring in interstitial lung diseases (ILD). In advance, we perform over-segmentation of the volumes to supervoxels of an average size of $1cm^3$. We rebag sets of supervoxels in a way to from training data equivalent to what would be training data from clinical source for the proposed algorithm. For our experiments, we extract two texture descriptors for each supervoxel: (1) 1200-dimensional Texture Bags on Local Binary Patterns (BVW) and (2) 52-dimensional Haralick features around the center of the supervoxel. For each descriptor, we generate a semantic profile mapping (SP-BVW and SP-Haralick). Based on a set of queries, we rank the training data using Euclidean distance among 4 feature vectors. Figure 1 (b) shows ground truth labelings of a volume, and figure 1 (b) a labeling obtained by assuming that the highest semantic profile coefficient is a good estimator for the correct label. Figure 2 shows precision-recall curves for the five tissue classes. The proposed method offers favourable runtime with less than three minutes for learning on 615000 instances and five classes. The calculation of the semantic profiles of 7526 descriptors (one lung) takes 1.37 seconds.

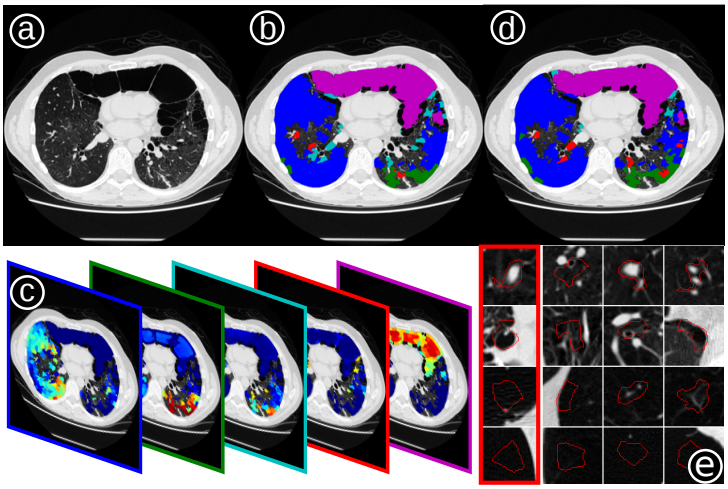

Figure 1: In medical imaging (a) only a small part of the information captured by visual features relates to relevant clinical information such as diseased tissue types (b). Information for learning is typically only available as sets of reported findings on the image level. We demonstrate how to learn a mapping of these weak semantic labels to individual voxels (c). This results in good labeling accuracy (d), and improved retrieval (d).

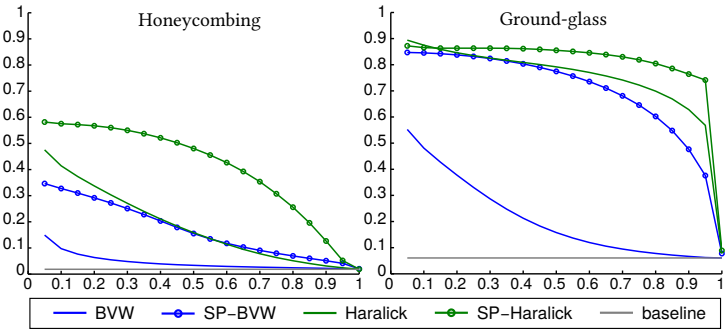

Figure 2: Precision Recall curves for two anomaly classes. Semantic Profiles (SP-BVW, SP-Haralick) consistently outperform the corresponding visual descriptors BVW and Haralick.

[1] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid. Multiple instance metric learning from automatically labeled bags of faces. In *Computer Vision–ECCV 2010*, pages 634–647. Springer, 2010.

[2] Rong Jin, Shijun Wang, and Zhi-Hua Zhou. Learning a distance metric from multi-instance multi-label data. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009.*, pages 896–902. IEEE, 2009.

[3] Mustafa Ozuysal, Pascal Fua, and Vincent Lepetit. Fast keypoint recognition in ten lines of code. In *Conference on Computer Vision and Pattern Recognition, CVPR 2007.*, pages 1–8. IEEE, 2007.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.