# Brain Tumor Region Segmentation using
# Local Co-occurrence Features and Conditional Random Fields

Siyamalan Manivannan[1], Haocheng Shen[1], Wenqi Li[1], Roberto Annunziata[1],
Hadi Hamad[2], Ruixuan Wang[1], and Jianguo Zhang[1]

[1] CVIP, School of Computing, University of Dundee, UK.
[2] Department of Mathematics and Information, University of Palermo.

**Abstract.** In this paper we propose a method to segment brain tumor regions in digital pathology images. Multiple local features as well as their local co-occurrences are used to characterise brain tissue regions. An SVM trained on these representations is used to predict the label of new tissue regions, followed by a conditional random field framework which can provide smoothed boundaries for predicted tumor regions. An average segmentation score of 0.66 was obtained based on a 5-fold cross validation.

## 1 Method

We propose a system to segment necrotic regions from normal regions in brain pathology images based on a sliding window classification followed by a conditional random field (CRF) smoothing.

### 1.1 Sliding window classification

We consider the tumor region segmentation as a window-based classification problem. Firstly a set of grid points with a step size of $M$ are defined on each image, then windows of size $W \times W$ centered at those grid points are extracted as shown in Figure 1. In the following, we will introduce how to represent image windows, and then how to train a SVM to classify new windows.



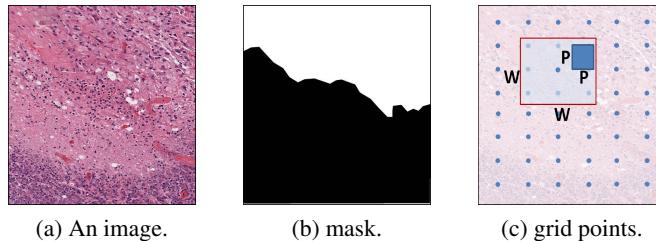(a) An image.          (b) mask.          (c) grid points.

Fig. 1: The generation of grid points and windows: (a) an image. (b) its corresponding mask, where black region indicates the presence of necrotic. (c) grid points with a step size of $M$ are defined on the images (blue circles) and image windows of size $W \times W$ centered at those grid points are extracted for processing (red square indicates an example window), and patches of size $P \times P$ inside each windows are considered for feature extraction (blue square).

**Feature extraction:** Each window is preprocessed before feature extraction, i.e., the pixel values in each channel of HSV color space are linearly rescaled so that 1% of pixels became saturated at low and high values respectively. Then image patches of size $P \times P$ ($P << W$), with an overlap of $\frac{2}{3}P$ pixels in both horizontal and vertical directions are extracted inside each window. The following four types of local features are used to describe each image patch:

1. *Multi-resolution local patterns (mLP)*: a 3-resolution version of non-binarized local binary pattern feature, recently proposed for Human Epithelial cell and specimen classification in [8,9]. This feature is extracted from each of the HSV color channel and concatenated to get a feature representation.

2. *Root-SIFT (rSIFT)*: a variant of widely used SIFT feature proposed in [10] produces better performance than SIFT for image matching and retrieval tasks. As for mLP, rSIFT is extracted from each color channel and concatenated to get a representation.
3. *Random Projections (RP)*: a simple method for dimensionality reduction, projecting each linearised raw patch vector of size $D' = P \times P \times 3$ (3 indicates the HSV color channels) to a much compressed space of dimension $D$ ($D << D'$) [11]. RP are successfully applied for texture image classification in various applications, e.g. [8,9,11].
4. *Intensity histograms (IH)*: a 256 dimensional intensity histogram feature extracted from each of the HSV color channels and then concatenated as a feature vector.

Since the features (rSIFT and IH) are extracted from different color channels, the final dimensionality will be large; therefore PCA was applied on rSIFT and IH features to reduce their dimensionality to 300.

**Feature encoding:** Bag-of-words (BoW), a widely-used feature encoding method, assumes that the local features extracted from images are independent of each other, and only counts the frequency of each visual word appearing in each image. Recently, an inter-cluster feature encoding method was proposed to additionally capture the co-occurrence of "visual words" within intermediate-scale image regions (e.g., with size of $64 \times 64$ pixels), and showed improved classification performance over the BoW method in some medical domains [12].

Here we followed the approach proposed in [12] to represent each image window. For each feature type, we build two dictionaries, one with a size of 1000 visual words to capture the statistics of local features and the other one being a rather small dictionary of size 100 to capture the intermediate-scale information. Both dictionaries are built based on $300,000$ randomly sampled local features. As proposed in [12], we use 20% of cluster (i.e., word) pairs from the smaller dictionary to capture the intermediate-scale features. Finally each image window is represented using this representation, leading to a feature dimensionality of 1950 for each feature type.

**Classification:** A SVM classifier can be trained to categorize each image window into either the necrotic or the normal class. To train the SVM, each image window is represented based on the above feature encoding method, and the class label of each image window was generated based on the ground-truth segmentation masks, i.e., if the percent of the number of pixels being necrotic within the image window is larger than a threshold (e.g., 75%), the image window will be considered necrotic. Otherwise, the window is considered normal.

Since SVM does not provide probabilistic outputs, we use the Platt scaling [13] to convert the SVM outputs into probability values. The Platt calibration method [13] maps any SVM output $f(\mathbf{x})$ with the range $[-\infty, +\infty]$ to a posterior probability $P$ with the range $[0,1]$ by a sigmoid function, i.e.,

$$P(y = 1 | f(\mathbf{x})) = \frac{1}{1 + \exp^{af(\mathbf{x})+b}} \tag{1}$$

where $P(\cdot)$ represents the probability of the image window (represented by its feature $\mathbf{x}$) being positive. We use a 3-fold cross validation on the training image set to generate a training data for sigmoid learning as suggested by Platt [13].

After a SVM is trained, it can be used to segment a new image. More specifically, image windows are extracted and then individually classified by the SVM. Since every image pixel will be covered by multiple image windows, the probability of each pixel being necrotic is obtained by averaging the probabilities of these relevant image windows being necrotic, therefore generating a final probability map of the image (see Figure 2 for an example).

## 1.2 Refinements with CRF

To obtain the final segmentation with smooth boundaries and to discard noisy isolated predictions, A CRF framework was applied as in [14]. CRF provides a natural way to incorporate pair-wise constraints, enforcing adjacent regions belonging to the same class. Let $G(S, E)$ be the adjacency graph of pixels, with each pixel corresponding to a node $s \in S$, and every edge $(s_i, s_j) \in E$ indicating the neighborhood relationship between two pixels $s_i$ and $s_j$. Let $P(\mathbf{c}|G; w)$ be the conditional probability of the set of class label assignments $\mathbf{c}$ to all pixels given the adjacency graph $G(S, E)$ and a weight $w$. Then CRF minimizes an energy of the form [14]

$$-log(P(\mathbf{c}|G; w)) = \sum_{s_i \in S} \psi(c_i | s_i) + w \sum_{(s_i, s_j) \in E} \Phi(c_i, c_j | s_i, s_j) \tag{2}$$

We directly use the probability outputs $P(c_i = 1|s_i)$ from the probability map provided by SVM to define the node ($\psi$) and edge potentials ($\Phi$):

$$\psi(c_i|s_i) = -log(P(c_i|s_i)) \tag{3}$$

$$\Phi(c_i,c_j|s_i,s_j) = \frac{|c_i-c_j|}{1+|P(c_i|s_i)-P(c_j|s_j)|} \tag{4}$$

where the weight $w$ in Eq. 2 represents the trade-off between the spatial regularization (edge-potential) and the confidence in the classification (node-potential). $w$ is learned based on the cross-validation in the training set. We use the public library for graph-optimization [15–17] for the label inference.

## 2  Experiments

We evaluate the method using the four different features and their combinations, with and without CRF based refinement. The training dataset provided by the MICCAI 2014 digital pathology challenge contains 35 images with varying sizes, ranging from 500 to 5000 pixels in each dimensions. As test set is currently not released, we evaluated the proposed method based on a 5-fold cross validation. The following segmentation score $r$ is used to measure the accuracy of the proposed method in segmentation,

$$r = \frac{A \cap B}{A \cup B}, \tag{5}$$

where $A$ and $B$ are the ground-truth and the predicted segmentations respectively for any image. For the necrotic images (containing both necrotic and normal regions), $r$ is the ratio between the number of necrotic pixels which are correctly predicted by the method and the number of pixels which are labelled as necrotic in either the predicted segmentation or the ground truth. For the non-necrotic images (only containing normal regions), $r$ is the ratio between the number of non-necrotic pixels which are correctly predicted by the method and the number of pixels which are labelled as non-necrotic in either the predicted segmentation or ground truth. We report the segmentation score by averaging the scores over all the images.

The step size for grid points ($M$) and the patch size ($P$) for local feature extraction are fixed to 75 pixels and 24 pixels respectively. We use the public library liblinear [18] for training the SVM. In the first two experiments CRF is not applied. The final segmentation is obtained by thresholding the averaged probability map of an image. The threshold was learned based on maximizing the segmentation score on the training data.

### 2.1  Performance of different features

This experiment evaluates the performance of different features and their combinations. mLP gives better performance compared to other individual features, and when it was combined with rSIFT, the combination (rSIFT+mLP) outperforms other individual features and combinations (Table 2). Local intensity histogram features perform worse compared to all other features. Therefore in all the following experiments, the combined features rSIFT+mLP are used. The window size in this experiment was fixed to 200.

| feature | dimensionality | score |
|---|---|---|
| mLP | 1990 | 0.6207 |
| rSIFT | 1990 | 0.5749 |
| RP | 1990 | 0.5853 |
| IH | 1990 | 0.4968 |
| mLP + rSIFT | 3980 | 0.6600 |
| rSIFT + RP | 3980 | 0.6247 |
| mLP + RP | 3980 | 0.6099 |
| mLP + rSIFT + RP | 5970 | 0.6515 |

Table 1: Performance of different features and their combinations.

| window size | score |
|---|---|
| 100 | 0.5702 |
| 200 | 0.6600 |
| 300 | 0.6597 |
| 400 | 0.6303 |
| 200 + 300 | 0.6660 |
| 100 + 200 + 300 + 400 | 0.6483 |

Table 2: Effect of window sizes.

## 2.2 The effect of window size

In this experiment we report the effect of different window sizes $W \times W$, $W \in \{100, 200, 300, 400\}$ and their combinations. The window size 200 gives better performance compared to other windows. When combining (averaging) the probability maps of windows of sizes 200 and 300, the segmentation score increases slightly. More specifically, two different SVM classifiers are learned, one for each window size. Then the probability maps obtained from window sizes 200 and 300 are averaged and compared with a learned threshold to obtain the final segmentation. This threshold was learned by maximizing the segmentation score on the training set as explained above.
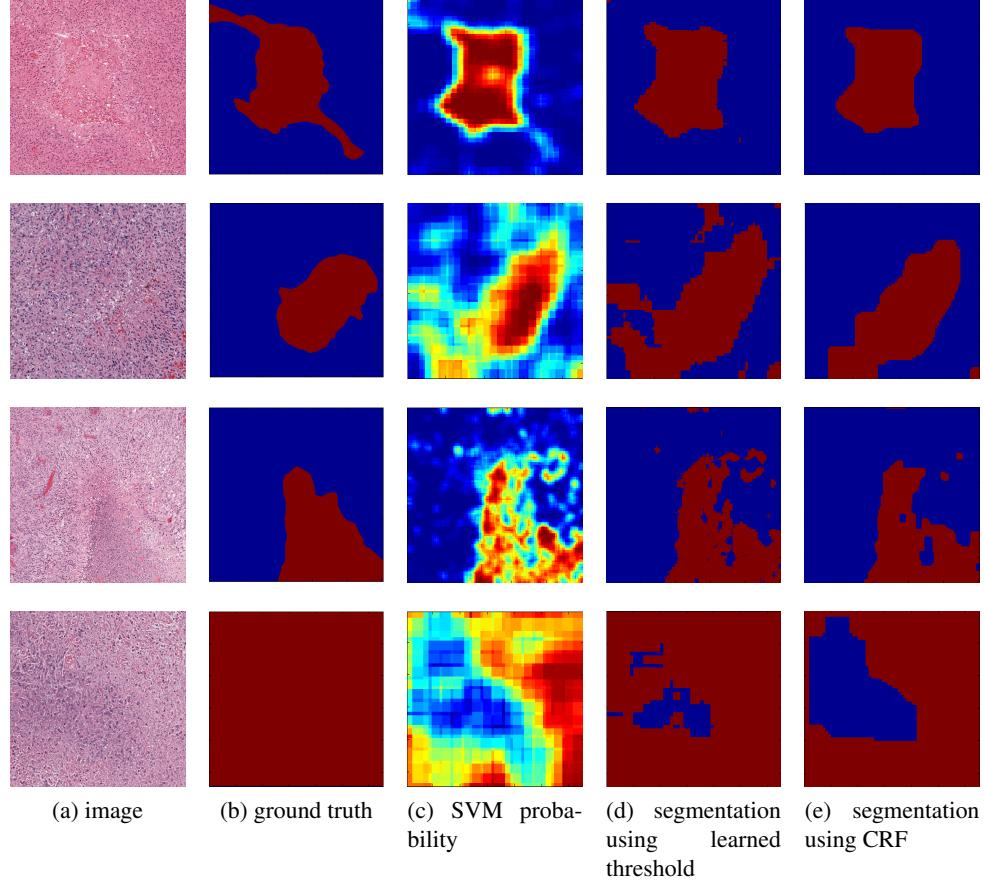


|(a) image | (b) ground truth | (c) SVM probability | (d) segmentation using learned threshold | (e) segmentation using CRF|

Fig. 2: Examplar segmentation results. (a) input images. (b) ground-truth segmentation. (c) the probability map based on SVM outputs. (d) the segmentation obtained by thresholding (c) with a learned threshold. (e) the segmentation obtained by CRF using (c) as the input. (red indicates the presence of necrotic, and blue indicates the presence of non-necrotic.)

## 2.3 The effect of CRF refinement

This experiment evaluates the performance of CRF. rSIFT+mLP features with window sizes of 200 and 300 pixels were used. The probability maps obtained from 200 and 300 window sizes are averaged and used as the input to the CRF framework. Figure 2 shows some outputs of the learned threshold based segmentation and the CRF based segmentation. As expected, CRF smooths the boundaries and removes most of the noisy predictions. But in this experiment we were unable to get any performance improvement in terms of the averaged segmentation score, probably because the performance of CRF completely relies on the SVM predictions. Better performance may be obtained by

adding raw features in addition to the SVM probabilities, which is currently explored for the 2nd phase of the MICCAI Challenge.

## 3    Conclusions

In conclusion, we proposed a region-based classification framework for the segmentation Challenge, based on recently developed local feature encoding and the conditional random field framework. Decent segmentation result has been obtained in average. Currently, we are still exploring to improve the performance by adding visual features into the CRF framework, and by employing more discriminative local features for the SVM training, particularly based on the deep learning approach.

## References

1. Demir, C., Yener, B.: Automated cancer diagnosis based on histopathological images: A systematic survey, Rensselaer Poly-technic Institute (2005)
2. Gurcan, M.N., Boucheron, L., Can, A., Madabhushi, A., Rajpoot, N., Yener, B.: Histopathological image analysis: A review. IEEE Reviews in Biomedical Engineering (2009)
3. Bhagavatula, R., Fickus, M., Kelly, W., Guo, C., Ozolek, J., Castro, C., Kovacevic, J.: Automatic identification and delineation of germ layer components in H and E stained images of teratomas derived from human and nonhuman primate embryonic stem cells. In: IEEE International Symposium on Biomedical Imaging. (2010)
4. Monaco, J., Tomaszewski, J., Feldman, M.D., Hagemann, I., Moradi, M., Mousavi, P., Boag, A., Davidson, C., Abolmaesumi, P., Madabhushi, A.: High-throughput detection of prostate cancer in histological sections using probabilistic pairwise markov models. Transactions on Medical Image Analysis (2010)
5. Tabesh, A., Teverovskiy, M., Pang, H.Y., Kumar, V., Verbel, D., Kotsianti, A., Saidi, O.: Multifeature prostate cancer diagnosis and gleason grading of histological images. IEEE Transactions on Medical Imaging (2007)
6. Nakazato, Y., Minami, Y., Kobayashi, H., Satomi, K., Anami, Y., Tsuta, K., Tanaka, R., Okada, M., Goya, T., Noguchi, M.: Nuclear grading of primary pulmonary adenocarcinomas. Cancer (2010)
7. Chang, H., Borowsky, A., Spellman, P., Parvin, B.: Classification of tumor histology via morphometric context. In: IEEE Conference on Computer Vision and Pattern Recognition. (2013)
8. Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., McKenna, S.J.: Hep-2 cell classification using multi-resolution local patterns and ensemble SVMs. In: ICPR I3A workshop on pattern recognition techniques for IIF images. (2014)
9. Manivannan, S., Li, W., Akbar, S., Wang, R., Zhang, J., McKenna, S.J.: Hep-2 specimen classification using multi-resolution local patterns and SVM. In: ICPR I3A workshop on pattern recognition techniques for IIF images. (2014)
10. Arandjelović, R., Zisserman, A.: Three things everyone should know to improve object retrieval. In: IEEE Computer Vision and Pattern Recognition. (2012)
11. Liu, L., Fieguth, P.: Texture classification from random features. IEEE Transactions on Pattern Analysis and Machine Intelligence (2012)
12. Manivannan, S., Wang, R., Trucco, E.: Inter-cluster features for medical image classification. In: MICCAI. (2014)
13. Platt, J.C.: Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In: Advances in large margin classifiers. (1999)
14. Fulkerson, B., Vedaldi, A., Soatto, S.: Class segmentation and object localization with superpixel neighborhoods. In: IEEE International Conference on Computer Vision. (2009)
15. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence (2004)
16. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence . (2004)
17. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence (2001)
18. Fan, R.E., Chang, K.W., Hsieh, C.J., Wang, X.R., Lin, C.J.: LIBLINEAR: A library for large linear classification. Journal of Machine Learning Research (2008)