Image Topic Discovery with Saliency Detection

Zhidong Li[1 2]
Zhidong.Li@nicta.com.au

Yang Wang[1 2]
Yang.Wang@nicta.com.au

Jing Chen[2]
Jng.Chen@ieee.org

Jie Xu[1 2]
Jxu@nicta.com.au

John Larid[1 2]
John.Larid@nicta.com.au

[1] The National ICT, AUSTRALIA

[2] The School of Computer Science and Engineering, University of the New South Wales, AUSTRALIA

Image topic discovery is a challenging task in computer vision, which is important for content understanding, image retrieval, and event detection. In recent years, image categorization by combination of Bag-of-Word (BoW) model and latent topic discovery models [1, 2] has gained considerable attention. Instead of reading word by word in a document, human is usually attracted by salient objects in the scne before noticing the other parts. In practice, a person can rapidly categorize the images into different classes based on the salient information captured with the prior knowledge [3] or merely through low-level visual attention.

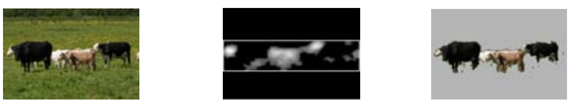
Inspired by the psychophysics study of human vision system, in this paper, we introduce a biologically inspired approach that combines latent topic model with saliency detection to category image dataset. First, based on PFT approach [4] and Grab-Cut, a saliency detection algorithm is proposed to discriminate salient objects from background parts in the image. Second, a hierarchical generative model is presented to discover image topics by considering subtopics of both salient objects and background parts in the image.

For the saliency detection, given an image I, we first compute its saliency map SM by Fourier transform F of the image and inverse Fourier transform F^{-1} [4]:

$$SM = g(I) * \left\| F^{-1}[e^{i \cdot P(F(I))}] \right\| \qquad (1)$$

Where $g(.)$ is a 2D Gaussian filter and $P(\cdot)$ is the image phase spectrum. The regions with saliency values higher than a threshold (0.75 times the maximum saliency value in the saliency map) are detected as salient objects [4], and the other areas become background parts.

However, the detected salient regions usually cannot represent the entire object. We handle the saliency activation spreading by Grab-Cut, which requires the user to input the rectangle area of foreground. To automatically detect the salient objects, we set the rectangle area as the bounding box of the salient regions (see figure 1b). Figure 1c shows the detection result by the proposed salient detection algorithm.



(a) (b) (c)

Figure 1: (a) One example image. (b) Result of saliency detection based on PFT approach. (c) Salient object detection by the proposed method.

A hierarchical generative model (see Figure 2) is used to integrate latent topic discovery with saliency detection. In this work, the model performs image topic discovery by combining subtopics of both salient objects and background parts in the image. Given a dataset of M images, two codebooks, one of visual words w^s from salient objects and the other of w^b visual words from background regions are formed through the K-means algorithm using SIFT feature. For each image I_m, latent topics t are sampled from a multinomial distribution with parameter θ, and θ is generated from a Dirichlet prior with hyper-parameter α. On the other hand, there are N_m^s saliency words and N_m^b background words observed in I_m. Each background word w^b is associated with one background subtopic z_k^b. Meanwhile each saliency word w^s is associated with a saliency subtopic z_k^s. Here z_k^s and z_k^b are sampled from the topic of image. The graph model is shown in figure 2.

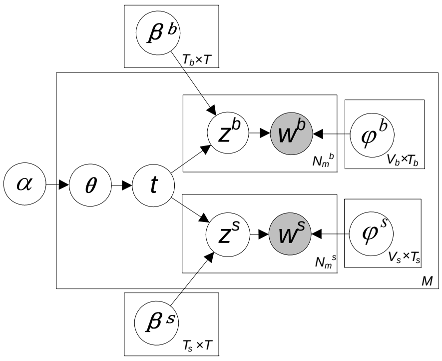


Figure 2: Proposed generative model of the topic discovery. The dark nodes represent observed data and the light nodes represent latent variables and parameters (see the text for details).

The experimental results show that the proposed approach robustly detects salient objects and categorizes image data, and it outperforms state-of-the-art methods for both saliency detection and unsupervised topic modelling.

	Accuracy
Proposed	56.30%
LDA[1]	51.37%

(a)

	Accuracy
Proposed	52.25%
Spatial_LTM[5]	48.25%

(b)

Figure 3: The overall accuracy of image categorization using MSRC and Corel based image dataset.

[1] D.M. Blei, A.Y. Ng and M.I. Jordan., Latent Dirichlet Allocation. Journal of Machine Learning Research, (3):993-1022. 2003.

[2] T. Hofmann. Probabilistic Latent Semantic Indexing. In Proc. SIGIR, 1999.

[3] P.V. Marius, F.F. Li and K. Sabine, Neural mechanisms of rapid natural scene categorization in human visual cortex. Nature 460: 94-97, 2009..

[4] C. Guo, Q. Ma and L. Zhang. Spatio-temporal Saliency Detection Using Phase Spectrum of Quaternion Fourier Transform. In Proc. CVPR. 2008.

[5] L. Cao and F.F. Li. Spatially coherent latent topic model for concurrent object segmentation and classification. In *Proc. ICCV*, 2007.