Anatomical Structure Sketcher for Cephalograms by Bimodal Deep Learning

Yuru Pei[1]
peiyuru@cis.pku.edu.cn
Bin Liu[1]
liubin@cis.pku.edu.cn
Hongbin Zha[1]
zha@cis.pku.edu.cn
Bing Han[2]
orthohanks@sina.com
Tianmin Xu[2]
tmxuortho@gmail.com

[1] Key Laboratory of Machine Perception (MOE)
Peking University
Beijing, China

[2] Stomatology Hospital
Peking University
Beijing, China

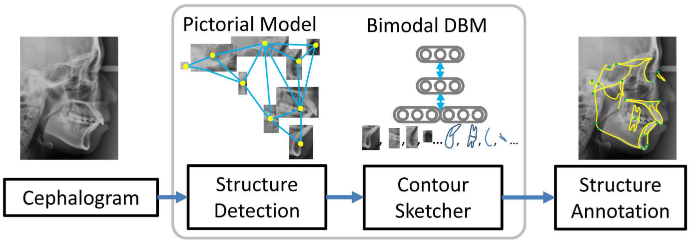

Figure 1: Flowchart of our cephalogram sketcher system.

Lateral cephalogram X-ray (LCX) images are essential to provide patient-specific morphological information of anatomical structures. The automatic annotation of anatomical structures in cephalograms has been performed in the biomedical engineering for nearly twenty years. Most systems only handle a portion of salient craniofacial landmark set [1, 2, 3]. Although model-based methods can produce a full set of markers [5, 7], the pattern fitting can fail to converge in blurry images. It is challenging to annotate LCX images with high fidelity.

In this work, we propose a novel cephalogram sketcher system as shown in Fig. 1 for the automatic anatomical-structure annotation, especially for the blemished images due to structure overlappings and device-specific distortions during projection. Firstly, we introduce an hierarchical extension of a pictorial model to detect anatomical structures. Secondly, the bimodal deep Boltzmann machine (DBM) is employed to sketch the structure contours. Specifically, the contour sketcher takes advantages of the path in the DBM to extract the contour definitions from the patch textures by alternating Gibbs sampling.

Given a cephalogram I, the structure definition S, and the parameters $\Theta = (\Theta_q, \Theta_r)$ with respect to the intra- and inter-layer correlations, the posterior probability distribution according to the Bayes rule is defined as $P(S|I,\Theta) \propto P(I|S,\Theta)P(S|\Theta)$, where $P(S|\Theta)$ is a shape prior distribution. $P(I|S,\Theta)$ is the image likelihood given the hierarchical architecture and the model parameters. The likelihood can be factorized as a product of likelihoods of local structures.

$$\begin{aligned}\ln P(I|S,\Theta_q,\Theta_r) = &\sum_{l=0}^{D-1} \sum_{s_i^l \in L_l} \ln\phi(s_i^l) + \\ &\sum_{l=0}^{D-2} \{\sum_{C_r} \ln\varphi(s_i^l, s_j^{l+1}|\Theta_{r,ij}) + \sum_{C_q} \ln\varphi(s_i^{l+1}, s_j^{l+1}|\Theta_{q,ij})\}.\end{aligned} \quad (1)$$

$\phi(s_i)$ is potential of local structures computed based on the output of SVM classifiers as in [4], and $\phi(s_i) = (1 + \exp(A_i f_i(s_i) + B_i))^{-1}$, where f_i is the output of the linear SVM classifier for the i_{th} kind of structures s_i. A linear SVM is trained for each kind of structures respectively, where the parameters, A_i and B_i, are predefined.

When given multimodal data, the deep architecture can build a joint representation by virtue of hidden layers from each modality. The shared attributes can be transferred from one modality to the other via the joint layer. Considering the undirected RBM, the observed data of one modality take a role in training the parameters related to layers of the other modality, which is the main difference from the multimodal DBN [6].

The bottom layers $(v_m, h_m^{(1)}, \ldots, h_m^{(K-1)} | m = p,t)$ are related to the image patches and contour shapes respectively, where K is the number of hidden layers and set at 3 in our experiments. The top one $h^{(K)}$ is the joint hidden layer connecting both modalities. Given two modalities, the patches v_p and shape contours v_t, the joint distribution

$$\begin{aligned}p(v_p, v_t|\vartheta) = &\sum_{h_p^{(K-1)}, h_t^{(K-1)}, h^{(K)}} \exp\left(-E_j\right) \cdot \\ &\prod_{m=p,t} \sum_{h_m} \left(\exp\left(-E_g\left(v_m, h_m^{(1)}\right) - \ldots - E_b\left(h_m^{(K-2)}, h_m^{(K-1)}\right)\right)\right).\end{aligned} \quad (2)$$

The partition functions are removed for clarity. The term E_j is the energy between the joint hidden layer $h^{(K)}$ and the upper hidden layers with respect to the patch and contour modalities, $h_p^{(K-1)}$ and $h_t^{(K-1)}$, and

$$E_j = -\sum_i^{N_{h,K}} a_i^K h_i^{(K)} - \sum_m \left(\sum_j^{N_{h,K-1}^m} a_{m,j}^{K-1} h_{m,j}^{(K-1)} + \sum_i^{N_{h,K-1}^m} \sum_j^{N_{h,K}} h_{m,i}^{(K-1)} W_{ij}^{K,m} h_j^{(K)} \right). \quad (3)$$

It is intractable to learn the parameters by maximizing the above likelihood. Alternatively, the problem can be solved by mean-fields, which minimized the KL-divergence $p(h|v_p, v_t)$ [6]. The model parameters are initialized by learning the layer-wise stacking of RBMs.

For the purpose of contour sketching, $p(v_t|v_p, \vartheta)$ needs to be solved. In DBM, when given the observed modality, v_p, the missing modality v_t can be generated by alternating Gibbs sampling. Specifically, v_p serves as input, while all the hidden units are initialized randomly (e.g. set at zeros in our case). In order to decide when to stop the iteration, we measure a score of the contour predication by the distance between the input image patches and the reconstructed ones.

$$P(v_t|v_p, \vartheta) \propto \exp\left(-\left\|v_p - v_p^{model}\right\|\right), \quad (4)$$

where v_p^{model} is the reconstructed image patch features, i.e. the HOG histograms in our system, and v_p is the input.

Implementation of this method is described in the paper. Our conclusion is that the sketcher as an integration of the hierarchical pictorial model and the deep learning can infer the positions and contours of anatomical structures effectively, and robust to noisy data.

[1] D. Giordano, R. Leonardi, F. Maiorana, G. Cristaldi, and M. Distefano. Automatic landmarking of cephalograms by cellular neural networks. *Artificial Intelligence in Medicine*, pages 333–342, 2005.

[2] A. Innes, V. Ciesielski, J. Mamutil, and S. John. Landmark detection for cephalometric radiology images using pulse coupled neural networks. In *Proc. Int. Conf. on Artificial Intelligence*, volume 2, 2003.

[3] R. Leonardi, D. Giordano, and F. Maiorana. An evaluation of cellular neural networks for the automatic identification of cephalometric landmarks on digital images. *Journal of Biomed and Biotec*, 2009.

[4] John Platt et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[5] AA Saad, A. El-Bialy, AH Kandil, and AA Sayed. Automatic cephalometric analysis using active appearance model and simulated annealing. *Int J on GVIP*, 6:51–67, 2006.

[6] Nitish Srivastava and Ruslan Salakhutdinov. Multimodal learning with deep boltzmann machines. In *Proc. NIPS*, 2012.

[7] W. Yue, D. Yin, C. Li, G. Wang, and T. Xu. Automated 2-d cephalometric analysis on x-ray images by a model-based approach. *IEEE Trans. on Biomedical Engineering,*, 53(8):1615–1623, 2006.