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

Abstract

The lateral cephalogram is a commonly used medium to acquire patient-specific morphology for diagnose and treatment planning in clinical dentistry. The robust anatomical structure detection and accurate annotation remain challenging considering the personal skeletal variations and image blurs caused by device-specific projection magnification, together with structure overlapping in the lateral cephalograms. We propose a novel cephalogram sketcher system, where the contour extraction of anatomical structures is formulated as a cross-modal morphology transfer from regular image patches to arbitrary curves. Specifically, the image patches of structures of interest are located by a hierarchical pictorial model. The automatic contour sketcher converts the image patch to a morphable boundary curve via a bimodal deep Boltzmann machine. The deep machine learns a joint representation of patch textures and contours, and forms a path from one modality (patches) to the other (contours). Thus, the sketcher can infer the contours by alternating Gibbs sampling along the path in a manner similar to the data completion. The proposed method is robust not only to structure detection, but also tends to produce accurate structure shapes and landmarks even in blurry X-ray images. The experiments performed on clinically captured cephalograms demonstrate the effectiveness of our method.

1 Introduction

Lateral cephalogram X-ray (LCX) images are essential to provide patient-specific morphological information of anatomical structures, such as teeth and craniofacial skeletons. The manual annotation of the anatomical structures requires the practitioner's experiences. The probably involved errors come from annotation variances of different practitioners and the device-specific distortions due to radiographic film magnification rates. Moreover, the LCX images can be blurry due to the overlapping of the left and right-sided craniofacial structures.

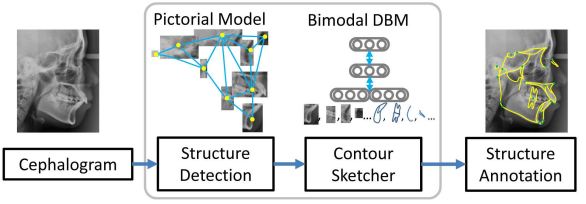


Figure 1: Flowchart of our cephalogram sketcher system.

The automatic annotation of anatomical structures in cephalograms has been performed in the biomedical engineering for nearly twenty years. From 1990s, the knowledge-based methods [8, 22] and the neural networks, e.g. pulse coupled neural networks [13] and cellular neural networks [9, 15], have been used in automatic annotation of cephalograms. Considering that most salient markers are on the boundary of projected structures in LCX images, various image filters including the canny filters [28] and mathematical morphological operators [2, 10], were introduced to the cephalogram analysis. The shape and appearance-model-based methods, e.g. the active shape model (ASM) [3] and active appearance model (AAM) [5] have been proved to be efficient and robust in facial feature extraction. The AAM and ASM have also been applied to the LCX images [23, 29]. Recently, the transfer learning technique was employed to get marker definitions from already annotated images instead of starting from scratch [27]. The existed work shows a nice performance on well-captured LCX images, where projected skeletal structures have clear boundaries. However, inherent blurs due to structure overlappings are unavoidable in the one-shot LCX images, and often blemish the feature extraction. Moreover, most systems only handle a portion of salient craniofacial landmark set [9, 13, 15]. Although model-based methods can produce a full set of markers [23, 29], the pattern fitting can fail to converge in blurry images. It is challenging to automatically annotate anatomical structures with high fidelity in LCX images.

In this work, we propose a novel cephalogram sketcher system as shown in Fig. 1 for the automatic annotation of anatomical structures, 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 is employed to sketch the structure contours. The deep machine tends to find a compact representation of observations, and the joint hidden layer can accommodate intrinsic structures from both modalities. Via the joint layer, a path between two modalities is established and the morphology data can be transferred. Specifically, the contour sketcher takes advantages of the path to extract the contour definitions from the patch textures by alternating Gibbs sampling. In the training process, the data of both modalities are present to learn a joint representation. In the sketching process, the contours are absent, and the inference can be seen as the completion of the missing data. To summarize, the contributions of this paper are:

- An application of the bimodal DBM to contour sketching and landmarking of anatomical structures in LCX images.
- An hierarchical extension of the pictorial model to detect craniofacial skeletal structures.
- A joint representation learning for dense textures of structure patches together with the sparse and arbitrary contours.

2 Related Work

Deep learning. The deep learning draws great interests in recent years. The elementary concepts and reviews can be found in [1, 4]. The deep learning provides a robust approach to find a compact representation of observed data. In the seminal work of deep learning [11], Hinton *et al*. used the stacking of several restricted Boltzmann machines (RBM) to learn the low-dimensional codes of face and digit images [14]. The deep learning has been widely used in the generative shape modeling [6, 12], the discriminative image classification [20], and the acoustic modeling for phone recognition [18].

In the deep structure, the hidden layer of one RBM serves as the observation of the higher RBM. However, the deep structure learnt layer-by-layer and followed by fine-tuning is different from the deep Boltzmann machine [24], where all the hidden layers are solved together with an undirected path between neighboring layers. Compared to a single layer of stochastic hidden units in the RBM and the deep energy machine [20], the computation of conditional posterior over multiple hidden layers is expensive and intractable [11, 24]. The contrastive wake sleep algorithm was used to learn multiple hidden layers of a deep belief network (DBN) [11]. An approximate inference was proposed to train DBM [25] efficiently, where the mean-fields techniques were used to minimize the KL-divergence of the approximated posteriors. The shape Boltzmann machine [6] subdivided the visible and the first hidden layers with overlappings for the purpose of efficient parameter learning.

Multimodal deep learning. The deep learning has been applied in multimodal learning under an assumption that different modalities of the same contents share a similar representation layer. The multimodal learning has very tantalizing properties that the missing modality can be filled in given the observed one. Even when some data from one modality is absent, the deep structure can still learn the joint representation of multi-modalities. Ngiam *et al*. [19] investigated the correlations between the audio and visual data by a shared hidden layer. All connections in the multimodal DBN were directional, and a shared representation layer connected with the hidden layers of the audio and visual modalities. Srivastava *et al*. [26] proposed the multimodal DBM for the text-image retrieval from unimodal and multimodal queries, which combined the text- and image-specific DBMs for a joint representation. The mean-field inference together with an MCMC-based stochastic approximation was used to estimate the model's stochastic parameters. The undirected DBM is more powerful than DBN considering the bi-way sampling during the learning process which can make use of both modalities for an optimal parameter selection. Luo *et al*. [16] employed the multimodal DBN to infer binary label maps from components patches for the purpose of facial feature segmentations. Different from the close regular regions in facial components parsing, the shapes of anatomical structures in LCX images are arbitrary and often with open contours. For instance, the contour of the lower brink of the cranium basion is complicated (see region 9 in Fig. 2). The binary maps in [16] can be no longer used to locate the anatomical structures and landmarks. In our work, we employ the parametric curves to represent shapes of anatomical structures. The bimodal DBM is employed to find a joint hidden layer of observed textures of image patches together with the sparse contour curves.

3 Anatomical Structure Detection

The cephalogram is an X-ray projection of the patient's head. The image can be parsed into a set of anatomical structures. The spatial configurations of anatomical structures vary

among individuals considering skeletal morphologies, head sizes, and subtle head movements during image capturing. As shown in Fig. 2(a)(b), the set of anatomical structures of interest is denoted as $S = \{s_i | i = 1, ..., 10\}$ including *nasion*, *orbitale*, *upper incisor and nasal spine*, *lower incisor and menton*, *upper molar*, *lower molar*, *porion*, *Pt*, *curve of sella and basion*, and *mandible*. The patches related to anatomical structures are represented by positions and sizes of the bounding boxes, and $s_i = (x_i^{anc}, \dot{w}_i, \dot{h}_i)$ with box anchor point x_i^{anc} . The predefined width and height are denoted as $\dot{w}_i$ and $\dot{h}_i$ respectively.

The structure detection can be performed by a pictorial model [7], which is under an assumption that all parts are nearly conditional independent given their neighbors. Thus, the likelihood probability of the holistic shape can be estimated by a product of local ones. However, considering the structure definition as shown in Fig. 2(a), most structures in L-CX images are overlapping, and some structures are even totally inside the others. For instance, the second and the eighth structures are inside the structure of *mandible*. There is an ad-hoc hierarchical architecture as shown in Fig. 2(c). Two kinds of correlations exist in the hierarchical architecture, the intra- and inter-layer correlations denoted as C_q and C_r respectively. As illustrated in Fig. 2(c), $C_q = \bigcup_{i=0}^{D-1} c_{i,i}$ are structure correlations inside the same layer L, and $C_r = \bigcup_{i=0}^{D-2} c_{i,i+1}$ related to correlations between layers, where $c_{l,l'} = \{< s_i, s_j > | s_i \in L_l, s_j \in L_{l'}\}$. D is the number of layers in the hierarchical architecture, and set at 3 in our experiments. The edge set in the graph is denoted as $C = C_q \cup C_r$.

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. In this work, the logarithm of the likelihood is rewritten as a combination of those related to the inter- and intra-layers (with partition functions removed for clarity).

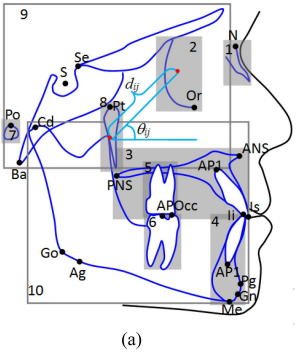
$$\begin{aligned} \ln P(I|S,\Theta_q,\Theta_r) = & \sum_{l=0}^{D-1} \sum_{s_i^l \in L_l} \ln \phi(s_i^l) + \\ & \sum_{l=0}^{D-2} \{ \sum_{<s_i^l, s_j^{l+1}> \in C_r} \ln \varphi(s_i^l, s_j^{l+1} | \Theta_{r,ij}) + \sum_{<s_i^{l+1}, s_j^{l+1}> \in C_q} \ln \varphi(s_i^{l+1}, s_j^{l+1} | \Theta_{q,ij})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

$\phi(s_i)$ is potential of local structures computed based on the output of SVM classifiers as in [21], 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. We employ the HOG to describe the texture feature of local patches. The image patch is subdivided in to 8×8 cells, and neighboring 2×2 cells are blocked. The histogram after gradient vector voting inside each cell is normalized by its surrounding four blocks containing this cell. Here the bin number is set at 9, and each cell is represented by a 36-dimensional vector. A linear SVM is trained for each kind of structures respectively, where the parameters, A_i and B_i, are predefined.

The pairwise potential $\varphi(s_i, s_j | \Theta_{ij})$ accounts for the correlations among neighboring structures. We consider two kinds of relationships, the distance d_{ij} between s_i and s_j, and the angle θ_{ij} defined by the line segment connecting the centers of two structures and the horizontal axis as shown in Fig. 2(a).

$$\varphi(s_i, s_j | \Theta_{ij}) = \mathcal{N}(d_{ij}, \theta_{ij} | \mu_{ij}, \Sigma_{ij}). \tag{2}$$

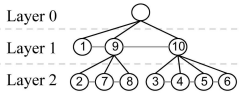
Given the training data, the Gaussian distribution parameters, (μ_{ij}, Σ_{ij}), can be obtained.



(a)

	Structures	Landmarks
1	Nasion	N
2	Orbitale	Or
3	Upper incisor and nasal spine	PNS, AP1, ANS, Is
4	Lower incisor and Menton	Ii, AP1, Pg, Me, Gn
5	Upper molar	APocc
6	Lower molar	APocc
7	Porion	Po
8	Pt	Pt
9	Curve of sella and basion	Ba, S, Se
10	Mandible	Go, Ag, Cd

(b)



(c)

Figure 2: The illustration of anatomical structures (a) and landmarks definitions (b). (c)The hierarchical architecture of ten anatomical structures.

The hierarchical pictorial model can locate anatomical structures and result to image patches containing structures by maximizing the likelihood (Eq.1).

4 Contour Sketcher

The contour-sketcher based on the bimodal deep learning is proposed for accurate structure contours and landmarks. The multimodal deep Boltzmann machine [26] is employed to build a joint representation of the image patch (one modality) and the corresponding contours (the other modality). In online sketching process, the unknown contours of the input patches are absent, and can be predicted by Gibbs sampling on the hidden modalities from the conditional distributions.

The contours $\mathbf{t}$ of the anatomical structures are discretized into a set of points, and $\mathbf{t} = (x_1, y_1, x_2, y_2, ..., x_n, y_n)$, where n is the number of points sampled on structures contours. The point sets of contours are predefined considering the sizes and shapes of different anatomical structures. The anchor point $t^{anc} = \left(\sum_{i=1}^{n} x_i/n, \sum_{i=1}^{n} y_i/n\right)$. The contours are aligned accordingly to the anchor point and $\tilde{\mathbf{t}} = \mathbf{t} - t^{anc}\mathbf{I}_{[1\times n]}$. All the contour anchors are moved to the origin. The anchor point together with the aligned contours $(t^{anc}, \tilde{\mathbf{t}})$ is used to represent the contour shapes of anatomical structures.

4.1 Deep Learning

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. There are symmetrically coupled stochastic units in DBM as a stacking of RBMs. Considering the undirected RBM, there is a complete path from one modality to the other. Specifically, 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 [26]. As shown in Fig. 3, one four-layer network is used to model the correlations between image patches and contour shapes. The bottom layers $(v_m, h_m^{(1)}, ..., h_m^{(K-1)} | m = p, t)$ 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 $h^{(K)}$

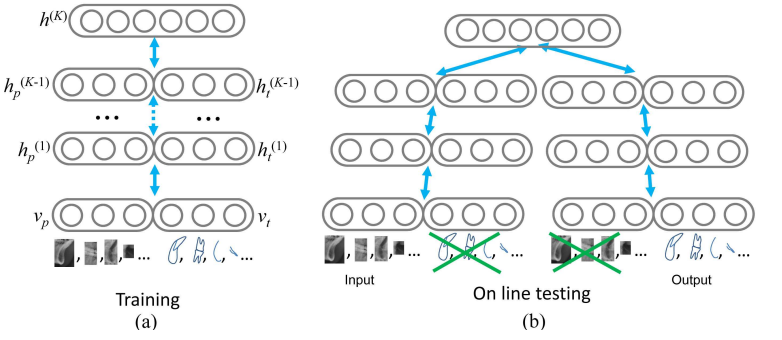


Figure 3: The training (a) and testing (b) of bimodal deep Boltzmann machine for contour sketching.

is the joint hidden layer connecting both modalities. In our work, two kinds of RBMs, the Gaussian-RBM (for visible-hidden layers) and Bernulli-RBM (for hidden-hidden layers), are used to build the deep architecture. The joint energies of visible and hidden units in Gaussian-RBM and Bernulli-RBM are defined as follows:

$$E_g(v,h^{(1)}|\vartheta) = \sum_i^{N_v}(v_i-b_i)^2\Big/2\sigma^2 - \sum_j^{N_{h,1}} a_j^1 h_j^{(1)} - \sum_i^{N_v}\sum_j^{N_{h,1}} v_i W_{ij}^1 h_j^{(1)}, \text{ and}$$

$$E_b(h^{(l)},h^{(l+1)}|\vartheta) = -\sum_i^{N_{h,l}} a_i^l h_i^{(l)} - \sum_j^{N_{h,l+1}} a_j^{l+1} h_j^{(l+1)} - \sum_i^{N_{h,l}}\sum_j^{N_{h,l+1}} h_i^l W_{ij}^{l+1} h_j^{(l+1)},$$

where N_v and N_h are the number of units in the visible and hidden layers. The parameters $\vartheta = (W,\ \sigma,\ a,\ b)$ include the mapping matrix W between adjacent layers and biases, a and b, within each layers. In our experiments, all input variables are normalized and the variance σ of the Gaussian distribution is set at 1. Given two modalities, the patches v_p and the shape contours v_t, the joint distribution

$$\begin{aligned} p(v_p,v_t|\vartheta) = & \sum_{h_p^{(K-1)},h_t^{(K-1)},\,h^{(K)}} \exp(-E_j)\cdot \\ & \prod_{m=p,t}\ \sum_{h_m^{(1)},\ \ldots,\ h_m^{(K-2)}} \left(\exp\left(-E_g\left(v_m,h_m^{(1)}\right) - \ldots - E_b\left(h_m^{(K-2)},h_m^{(K-1)}\right)\right)\right). \end{aligned} \tag{3}$$

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

$$E_j = -\sum_i^{N_{h,K}} a_i^K h_i^{(K)} - \sum_{m=p,t}\left(\sum_j^{N_{h,K-1}^m} a_{m,j}^{K-1} h_{m,j}^{(K-1)} + \sum_i^{N_{h,K-1}^m}\sum_j^{N_{h,K}} h_{m,i}^{(K-1)} W_{ij}^{K,m} h_j^{(K)}\right). \tag{4}$$

It is intractable to learn the parameters by maximizing the above likelihood. Alternatively, the problem can be solved by variational approximation techniques, mean-fields, to minimize the KL-divergence between the approximated and the true posteriors $p(h|v_p,v_t)$ [26]. The model parameters are initialized by learning the layer-wise stacking of RBMs.

For the purpose of contour sketching, the conditional probability $p(v_t|v_p,\vartheta)$ 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). The contours are absent from input, and can be generated when go through the joint layers from the patch machine to the contour machine. 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.

$$P(v_t|v_p,\vartheta) \propto \exp\left(-\left\|v_p - v_p^{model}\right\|\right), \tag{5}$$

where v_p^{model} is the reconstructed image patch features, i.e. the HOG histograms in our system, and v_p the input. The inference is performed under an assumption that the more similar the estimated patches textures to the input, the more reasonable contours resulted.

4.2 Structure Inference: Positions and Contours

In order to acquire the positions and contours of anatomical structures, we employ a greedy searching technique as summarized in Algorithm 1. To begin with, the positions of all structures are initialized by the mean values of the Gaussian distributions of the inter- and intra-layer correlations (Eq. 2). And then, for each structure, the candidate positions with high classifier scores within the variances of the Gaussian distribution (Eq. 2) are explored, while the shape parameters of all other structures are held fixed. The log-likelihood (Eq. 1) is computed as the detection score. The conditional probability of the sketcher (Eq. 5) is computed as the sketching scores. The shape and contour variables with the high detection and sketching scores are assigned to the current structure. All other structures are processed in the same way. The iterative process continues until the combined scores of all anatomical structures are lower than the predefined threshold η or the iteration number is large enough.

Algorithm 1 LCX-Image-Sketcher

Input: LCX images, score threshold η, and *Iter_MAX*;
Output: Contours and landmarks of anatomical structures;

Initialize all structure positions by mean values of the correlation distribution (Eq. 2);
Evaluate individual structure classifiers by the linear SVM (Section 3);
$Iter\_num = 0$;
while *Iter_num*< *Iter_MAX* **do**
  for Each kind of anatomical structures **do**
    Explore candidate locations with high classifier scores within correlation variances of $\mathcal{N}\left(d_{ij},\theta_{i,j}\right)$ (Eq. 2);
    Compute detection scores as Eq. 1;
    Compute sketching score as Eq. 5;
    Assign shape and contour variables with high detection and sketching scores ($\geq \eta$) to the current structure;
  end for
end while

Table 1: The contours errors (mm^2) and landmarks errors (mm) of our methods compared with the AAM, the extended ASM (EASM), and the multimodal DBN (MDBN).

Contour	$s1$	$s2$	$s3$	$s4$	$s5$	$s6$	$s7$	$s8$	$s9$	$s10$	avg.
Ours	**25.3**	**34.3**	**30.6**	**17.9**	**37.7**	**49.7**	**28.1**	**26.7**	**34.0**	48.4	**33.3**
AAM	37.1	41.7	32.7	31.5	52.6	57.2	59.5	35.7	44.4	**44.2**	43.7
EASM	49.9	37.0	36.0	93.7	57.3	69.9	64.0	36.3	35.2	53.5	53.3
MDBN	72.5	n/a	57.2	104	91.5	105	51.2	60.0	n/a	n/a	77.5
Marker	Se	Or	N	Ba	Cd	PNS	Ls	Li	Me	Go	avg.
Ours	**0.345**	1.04	**0.315**	**0.373**	**1.16**	**0.670**	**0.936**	**1.05**	**1.13**	**0.655**	**0.768**
AAM	0.806	2.10	1.16	0.815	1.24	1.93	1.05	1.61	1.96	0.900	1.36
EASM	0.652	**0.973**	1.68	1.42	1.55	1.31	1.79	1.36	1.24	1.73	1.37
MDBN	n/a	n/a	2.70	n/a	n/a	2.11	2.67	1.27	n/a	n/a	2.18

5 Experiments

Data Set. The data set includes 724 LCX images captured in the clinical dental hospital. The resolution of digitalized LCX images is 720×900. In our system, 60% images are used as the training data, and the remaining for testing. The training data are manually annotated by two practitioners for bounding boxes and contours of the local anatomical structures. The bounding boxes are used to train the structure detectors, and the contour shapes are used to learn the sketcher. There exist blurs in most images of the data set due to structure overlappings and sometimes subtle movements during capturing.

Methods. We compare our sketcher method with the mainstream techniques in craniofacial components detection and feature extraction, including those using the AAM [5], the extended ASM [17], and the multimodal DBN [16].

5.1 Detection

Ten kinds of structures are considered in our experiments. The annotated bounding boxes in the training LCX images are used to learn the hierarchical pictorial model. The structure detection results are shown in the first row in Fig. 4 (a). The location error e^l is measured as a ratio of anchor point difference between the ground truth $x_{i,GT}^{anc}$ and the automatically detected x_i^{anc} to the size of bounding box. $e_{i,i}^l = \left\| x_{i,GT}^{anc} - x_i^{anc} \right\| \Big/ \sqrt{\dot{w}_i \dot{h}_i}$. The confusion matrix of location errors of ten anatomical structures is shown in Fig. 4 (b). The average error is 7.46e−2.

5.2 Sketcher

The bimodal DBM is trained with respect to each kind of anatomical structures. When given the image patches of the testing cephalogram, the related contour shapes can be inferred by the deep machine. In our experiments, the maximum number of iterations in Algorithm 1 is set at 30. The sketching results are shown in Fig. 4 side by side with those by the AAM [5], the extended ASM [17], and the multimodal DBN [16]. We follow the configurations of the MDBN in [16]. Only seven closed structures are located in the binary label maps by the MDBN as shown in the last row of Fig. 4 (a).

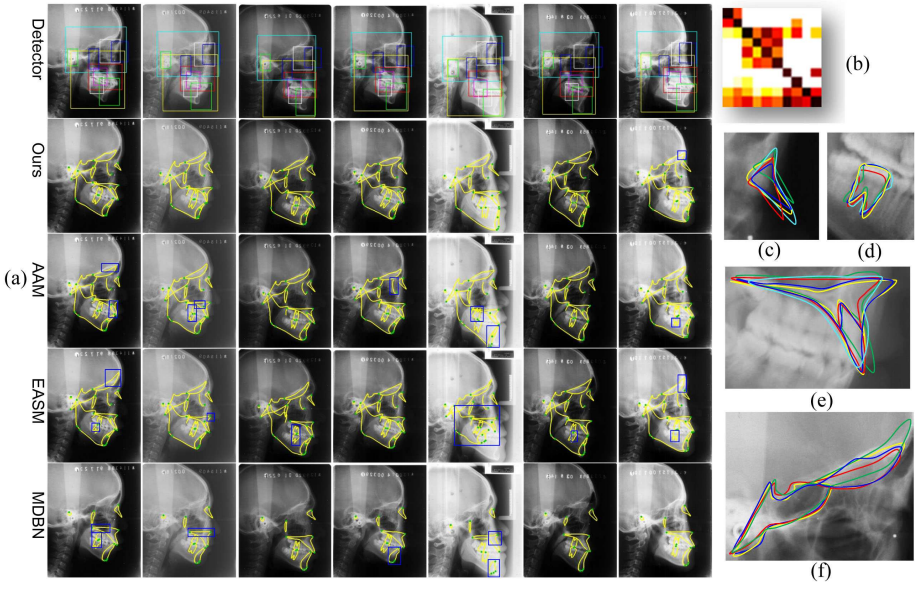


Figure 4: (a) The sketching results (yellow contours with green landmarks) of our methods and the AAM, the extended ASM, and the multimodal DBN (with failed cases blue outlined). (b)The confusion matrix of error e^l related to ten kinds of structures. The overlapping of contours extracted by all four methods and the ground truth of *nasion* (c), *lower molar* (d), *upper incisor* and *nasal spline* (e), and *curve of sella and basion* (f) (yellow-Ground truth, blue-Ours, red-AAM, green-EASM, cyan-MDBN).

There are inherent blurs in LCX images due to various reasons in clinical data capturing. The feature extractions based on the appearances and the principal shape models sometimes fail due to the image quality (See blue outlined in Fig. 4 (a)). For example, the extracted molars tend to shirk or swell, and even translate due to obscure images resulted from the overlapped neighboring teeth of the left- and right-sides. Moreover, it's relatively hard to cover the shape variations outside the training set. For instance, the aspect ratio of the craniofacial structure in the fifth case in Fig. 4 (a) is a bit different from others, and no such cases exist in the training set. The shape and appearance-based models fail to converge. The performance of the multimodal DBN seems to be more robust considering its deep structure. However, MDBN still fail to find the contour of *lower menton* of the fifth case. Our method can produce reasonable results even when all other methods fail due to the poor image quality. We also show the overlapping of contours extracted by all four methods and the ground truth of one LCX image in Fig. 4 (c-f) of *nasion*, *lower molar*, *upper incisor* and *nasal spline*, and *curve of sella and basion*. The contours computed by our method are closer to the real annotated ones.

The quantitative comparisons of all methods are illustrated in Table 1, where the errors are computed as the enveloped areas between the ground truth contours and automatically estimated. The average errors of our methods are lower than all other methods.

Landmarks. Some landmarks on the anatomical structures are important to clinical dentistry, especially to the orthodontics. There are heuristic rules [2] to locate landmarks when

given the contour shapes. For instance, *Or* marker is on the lower end of *Orbitale* contour. It's deserved to note that, in our system there is no explicit marker patterns like those in the AAM and ASM. The landmarks are all located by the heuristic rules as shown in Fig. 4 (a). Table 1 shows the errors computed as the L2 distance between the automatically located markers and the manually annotated.

6 Conclusions

In the paper, we propose a novel contour sketcher system for the LCX images, especially those blurred due to device-specific distortions or subtle head movements. By virtue of the bimodal DBM, the sketching problem is formulated as a cross-modal morphology transfer from texture features of regular image patches to arbitrary contour curves of structures. The sketcher as an integration of the hierarchical pictorial model and the deep learning can infer the positions and contours of anatomical structures effectively. The proposed method is robust to noisy data compared to state-of-the-arts.

Acknowledgement

This work was supported by NSFC 61272342, NHTRDP 863 Grant No.2012AA011602, NBRPC 973 Grant No. 2011CB302202.

References

[1] Yoshua Bengio. Learning deep architectures for ai. *Foundations and Trends in Machine Learning*, 2(1):1–127, 2009.

[2] J. Cardillo and MA Sid-Ahmed. An image processing system for locating craniofacial landmarks. *IEEE Trans. on Medical Imaging*, 13(2):275–289, 1994.

[3] Timothy F Cootes, Christopher J Taylor, David H Cooper, Jim Graham, et al. Active shape models-their training and application. *Computer vision and image understanding*, 61(1):38–59, 1995.

[4] Li Deng. An overview of deep-structured learning for information processing. In *Proc. Asian-Pacific Signal and Information Processing-Annual Summit and Conference (APSIPA-ASC)*, 2011.

[5] Gareth J Edwards, Christopher J Taylor, and Timothy F Cootes. Interpreting face images using active appearance models. In *Proc. Automatic Face and Gesture Recognition'98*, pages 300–305, 1998.

[6] SM Ali Eslami, Nicolas Heess, and John Winn. The shape boltzmann machine: a strong model of object shape. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 406–413, 2012.

[7] Martin A Fischler and Robert A Elschlager. The representation and matching of pictorial structures. *IEEE Transactions on Computers*, 100(1):67–92, 1973.

[8] DB Forsyth and DN Davis. Assessment of an automated cephalometric analysis system. *The European Journal of Orthodontics*, 18(5):471–478, 1996.

[9] 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.

[10] V. Grau, M. Alcaniz, MC Juan, C. Monserrat, and C. Knoll. Automatic localization of cephalometric landmarks. *Journal of Biomedical Informatics*, 34(3):146–156, 2001.

[11] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.

[12] Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh. A fast learning algorithm for deep belief nets. *Neural computation*, 18(7):1527–1554, 2006.

[13] 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.

[14] Honglak Lee, Chaitanya Ekanadham, and Andrew Ng. Sparse deep belief net model for visual area v2. *Advances in neural information processing systems*, 20:873–880, 2008.

[15] 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 Biomedicine and Biotechnology*, 2009.

[16] P. Luo, X. Wang, and X. Tang. Hierarchical face parsing via deep learning. In *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 2480–2487. IEEE, 2012.

[17] S. Milborrow and F. Nicolls. Locating facial features with an extended active shape model. *ECCV*, 2008. http://www.milbo.users.sonic.net/stasm.

[18] Abdel-rahman Mohamed, George E Dahl, and Geoffrey Hinton. Acoustic modeling using deep belief networks. *IEEE Trans. Audio, Speech, and Language Processing*, 20 (1):14–22, 2012.

[19] Jiquan Ngiam, Aditya Khosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y Ng. Multimodal deep learning. In *Proceedings of ICML'11*, pages 689–696, 2011.

[20] Jiquan Ngiam, Zhenghao Chen, Pang Wei Koh, and Andrew Y Ng. Learning deep energy models. In *Proc. of the 28th International Conference on Machine Learning*, volume 11, pages 1105–1112, 2012.

[21] 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.

[22] J. Ren, D. Liu, D. Feng, J. Shao, R. Zhao, Y. Liao, and Z. Lin. A knowledge-based automatic cephalometric analysis method. In *Proc. IEEE Engineering in Medicine and Biology Society*, volume 2, pages 723–727, 1998.

[23] AA Saad, A. El-Bialy, AH Kandil, and AA Sayed. Automatic cephalometric analysis using active appearance model and simulated annealing. *Int J on Graphics, Vision and Image Processing, Special Issue on Image Retrieval and Representation*, 6:51–67, 2006.

[24] Ruslan Salakhutdinov and Geoffrey E Hinton. Deep boltzmann machines. In *Proc. the international conference on artificial intelligence and statistics*, volume 5, pages 448–455. MIT Press Cambridge, MA, 2009.

[25] Ruslan Salakhutdinov and Hugo Larochelle. Efficient learning of deep boltzmann machines. In *Proc. Int. Conf. on Artificial Intelligence and Statistics*, 2010.

[26] Nitish Srivastava and Ruslan Salakhutdinov. Multimodal learning with deep boltzmann machines. In *Proc. Advances in Neural Information Processing Systems 25*, pages 2231–2239, 2012.

[27] WK Tam and HJ Lee. Improving point registration in dental cephalograms by two-stage rectified point translation transform. In *Proc. of SPIE*, volume 8314, 2012.

[28] J. Yang, X. Ling, Y. Lu, M. Wei, and G. Ding. Cephalometric image analysis and measurement for orthognathic surgery. *Medical and Biological Engineering and Computing*, 39(3):279–284, 2001.

[29] 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.