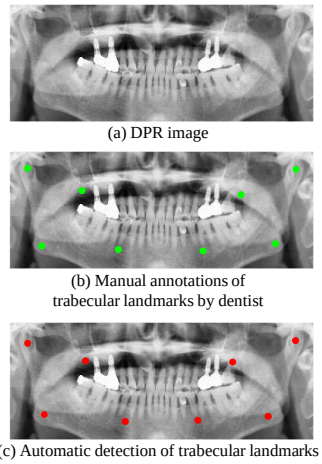
Detection of Trabecular Landmarks for Osteoporosis Prescreening in Dental Panoramic Radiographs

Jiaxiang Ren[1], Heng Fan[1], Jie Yang[2], and Haibin Ling[1]

***Abstract*— Dental panoramic radiography (DPR) images have recently attracted increasing attention in osteoporosis analysis because of their inner correlation. Many approaches leverage machine learning techniques (*e.g.* , deep convolutional neural networks (CNNs)) to study DPR images of a patient to provide initial analysis of osteoporosis, which demonstrates promising results and significantly reduces financial cost. However, these methods heavily rely on the trabecula landmarks of DPR images that requires a large amount of manual annotations by dentist, and thus are limited in practical application. Addressing this issue, we propose to automatically detect trabecular landmarks in DPR images. In specific, we first apply CNNs-based detector for trabecular landmark detection and analyze its limitations. Using CNNs-based detection as a baseline, we then introduce a statistic shape model (SSM) for trabecular landmark detection by taking advantage of spatial distribution prior of trabecular landmarks in DPR images and their structural relations. In experiment on 108 images, our solution outperforms CNNs-based detector. Moreover, compared to CNN-based detectors, our method avoids the needs of vast training samples, which is more practical in application.**

Clinical relevance **This paper presents an automatic way to detect jointly multiple trabecular landmarks from dental images for using in osteoporosis analysis.**



(a) DPR image

(b) Manual annotations of trabecular landmarks by dentist

(c) Automatic detection of trabecular landmarks by our solution

Fig. 1: Trabecular landmarks in a DPR image (a) of manual annotation by a dentist (b) and by our solution (c). We observe that, our method provides accurate localization of landmarks. Best viewed in color.

I. INTRODUCTION

Osteoporosis has been one of the most common diseases in US and affected over 3 million patients every year [26]. Numerous efforts have been made in recent years to provide early diagnosis of osteoporosis to reduce bone fracture risks. Currently, a golden standard is to utilize the Dual-energy X-ray Absorptiometry (DXA) for osteoporosis diagnosis [26], which is, however, cost inefficient and impractical for routine examination.

Inspired by the study of correlation between DXA measurements and dental data [9], [24], [12], many researchers analyze dental images to offer initial diagnosis of osteoporosis and demonstrate promising results. Among various types of image data, dental panoramic radiography (DPR) images have drawn extensive attention owing to its availability and promising results for the potential of osteoporosis prescreening. Kavita *et al*. [10] develop a support vector machine (SVM) learning method to identify osteoporosis or low bone mineral density. Roberts *et al*. [21] apply texture features extracted from DPR images for osteoporosis classification and achieve satisfied results. Li *et al*. [13] propose to classify osteoporosis condition with features from eight regions of interest (ROIs) that are centered at trabecular landmarks in DPR images. In order to improve classification accuracy, Bo *et al*. [4] propose a two stage SVMs model to classify osteoporosis condition with hand-crafted features from these RoIs. Furthermore, motivated by the powerful representation of deep convolutional neural networks (CNNs) [11], Chu *et al*. [6] propose an octuplet Siamese network (OSN) to learn discriminative features for osteoporosis status prediction. Yu *et al*. [27] propose a multi-task scheme for transfer learning in DPR images which achieves state-of-the-art performance.

Despite having achieved some promising results for osteoporosis prescreening, the aforementioned approaches heavily rely on the manual annotations of trabecular landmarks in DPR images for feature extraction (see Figure 1(a)), which severely limits its practical application. In this work, we aim to address this problem by proposing automatic detection of trabecular landmarks (see Figure 1(b)).

Motivated by the huge success of CNNs in generic object detection [22], [5], [20], [19], [15], [8], one natural solution is to directly apply existing state-of-the-art deep detector (*i.e.*, RetinaNet [14] to localize these trabecular landmarks. Nevertheless, we observe that, the CNNs-based detector fails to separate targets with similar intensity and texture features. Therefore, we use the CNNs-based detector as a baseline in this paper, and present a statistic shape model (SSM) for trabecular landmark detection. Our method is inspired by

*This work was supported in part by US NSF Grant 2002434

[1]Department of Computer Science, Stony Brook University, Stony Brook, NY, USA {jiaxren,hefan,hling}@cs.stonybrook.edu

[2]Division of Oral and Maxillofacial Radiology, School of Dentistry, Temple University, Philadelphia, PA, USA jie.yang@temple.edu

two observations in DPR images: (1) the relative positions of landmarks, on the left or right sides, are similar for each sample; (2) each type of landmarks lie in the similar position of a DPR image. For example, the condyle landmarks lie on the top corners of images. With such observations, we are able to use SSM to make full use of both the relationship and spatial prior distribution among landmarks. SSM is a statistical deformable model of the shape and appearance for a deformable class, which is then used as prior knowledge to locate the coordinates of landmark. In experiment on 108 images, our solution outperforms CNNs-based detector. Moreover, compared to CNN-based detectors, our method avoids the needs of vast training samples, which is more practical in application.

II. METHODOLOGY

A. Problem Formulation

We utilize a dataset of 108 DPR images extracted from different subjects for trabecular landmark detection [4], [6]. The images comes from 52 subjects with osteoporosis and 56 normal subjects. For each DPR image, eight ROIs correlated with the osteoporosis condition are manually annotated by dentists. Then our task is to localize these ROIs automatically, *e.g.* the x and y coordinates of ROIs in a DPR image. For simplicity, the trabecular ROIs are interchangeable with the trabecular landmarks in this task. For the i-th subject, it is expressed as

$$\boldsymbol{s}_i = \mathcal{G}(\boldsymbol{I}_i), \tag{1}$$

where $\boldsymbol{s}_i = [x_1, y_1, \ldots, x_{N_R}, y_{N_R}] \in \mathbb{R}^{2N_R \times 1}$ denotes the coordinates of ROIs, N_R (set to 8 in our work) represents the number of ROIs, $\boldsymbol{I}_i \in \mathbb{R}^{W \times H}$ is the input DPR image, and $\mathcal{G}$ the mapping function to be learned.

B. Deep Learning Approach

For a practical osteoporosis prescreening system, the CNNs-based detector should give consideration to both accuracy and efficiency. Inspired by the effectiveness of RetinaNet [14], we firstly introduce this detector to find ROIs in DPR images. The whole model can be divided into two parts, the deep feature extractor and the detection head. We freeze the feature extractor and only fine-tune the head to prevent overfitting. We also add a smaller anchor of 16×16 with stride of 4 to capture tiny texture features in trabecular. Considering the fact that trabecular patterns are scattered in various positions in the oral cavity, DPR images contain context information of trabecular bone. Thus the eight ROIs are grouped into four pairs based on symmetry of human face as illustrated in Figure 2. Each pair ROIs shares the same label, *e.g.* condyle, maxillary tuberosity, mandibular premolar region and mandibular angle.

During training, the targets are eight bounding boxes centered around the coordinates of the eight ROIs. Back-propagation is employed to update the weights in detection head. During inference, the outputs of detection network are bounding boxes with predicted labels. Here we propose two

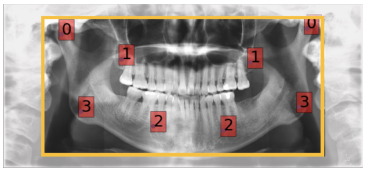


Fig. 2: Initial search box of SSM. Label-0: the condyle; Label-1: the maxillary tuberosity; Label-2: the mandibular premolar region; Label-3: the mandibular angle.

approaches to convert bounding boxes into the coordinates of ROIs $\boldsymbol{s}$: Global detection (GD) and local detection (LD). In GD, the conversion is produced in the left and right side respectively. Take the left side for example, the coordinates of each ROI (label) are the center of the bounding box with the highest confidence score for that label. Meanwhile the conversion in LD is produced only in the designated range where all training samples with the same ROI label lie within. Specifically, we firstly scan training set to find the minimum and maximum positions of the designated range. The dilation operation are adopted for more robust output. After that, only the bounding boxes that lie within the range are selected as the candidate bounding boxes. Finally, the coordinates of each ROI (label) are the center of the candidate box with the highest score for that label.

C. Landmark Localization Approach

In order to exploit the prior of spatial distribution of trabecular landmarks and their structural relationships, we propose using a statistic shape model (SSM) [7], [18] to detect DPR landmarks. SSM is a statistical deformable model of the shape and appearance for a deformable object class. A deformable object is generally taken as a set of landmarks points with specific semantic meanings. Using each DPR image and the 8 ROIs as a deformable object, SSM locates the coordinates of ROIs through optimization. In our task, a shape instance is interchangeable with a set of trabecular landmarks and represented as a $2N_R \times 1$ vector $\boldsymbol{s}$.

Given N training DPR images $[\mathcal{I}_1, \ldots, \mathcal{I}_N] \in \mathbb{R}^{WH \times N}$ with corresponding shapes $[\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N] \in \mathbb{R}^{2N_R \times N}$, the SSM consists several procedures. First, these training shapes are registered using generalized procrustes analysis (GPA) and then principal component analysis (PCA) is used to construct the orthonormal shape basis which is further augmented with four eigenvectors for the similarity transform. Mathematically, we reconstruct the i-th shape $\boldsymbol{s}_i$ as

$$\boldsymbol{s}_i = \bar{\boldsymbol{s}} + \boldsymbol{U}_s \boldsymbol{p}_i, \tag{2}$$

where $\bar{\boldsymbol{s}} \in \mathbb{R}^{2N_R \times 1}$ represents the mean shape in training set, $\boldsymbol{U}_s = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_n] \in \mathbb{R}^{2N_R \times n}$ is the augmented orthonormal shape basis, and $\boldsymbol{p}_i \in \mathbb{R}^{n \times 1}$ denotes the parameters of the i-th shape.

Afterwards, to take advantage of texture and intensity features in DPR images, we also construct an appearance model. Each training DPR image is wrapped into the mean shape and generates the shape normalized DPR image,

denoted as $\mathcal{I}_i(\mathcal{W}(\boldsymbol{p}_i))$, where $\mathcal{W}(\cdot)$ indicates the wrapping function. Thereafter, PCA is applied on $\{\mathcal{F}(\mathcal{I}_i)(\mathcal{W}(\boldsymbol{p}_i))\}_{i=1}^{N}$ to construct the orthonormal appearance basis, where $\mathcal{F}(\cdot)$ is the feature function such as LBP [1] or IGO [23]. Finally we can reconstruct the appearance instance of the i-th shape normalized DPR image $\boldsymbol{a}_i$ as

$$\boldsymbol{a}_i = \bar{\boldsymbol{a}} + \boldsymbol{U}_a \boldsymbol{c}_i, \tag{3}$$

where $\bar{\boldsymbol{a}} \in \mathbb{R}^{M\times 1}$ represents the mean feature of the shape normalized DPR images in training set, $\boldsymbol{U}_a = [\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_n] \in \mathbb{R}^{M\times m}$ denotes the orthonormal appearance basis, and $\boldsymbol{c}_i \in \mathbb{R}^{m\times 1}$ is the i-th appearance parameters.

The fitting procedure tries to recover parameterized description of a test sample through optimizing the objective function with respect to the shape and appearance parameters as follows

$$\arg\min_{\boldsymbol{p},\boldsymbol{c}} ||\mathcal{F}(\mathcal{I})(\mathcal{W}(\boldsymbol{p})) - \bar{\boldsymbol{a}} - \boldsymbol{U}_a \boldsymbol{c}||^2. \tag{4}$$

Lucas-Kanade Optimization [17], [3] is applied to solve the optimization in Eq. (4). With the best fitting parameters and the orthonormal basis of shape model $\boldsymbol{U}_s$, we can determine the test sample landmark $\hat{\boldsymbol{s}}$.

III. EXPERIMENTS AND RESULTS

In this section, we conduct experiments on the DPR dataset [4], [6] to validate the effectiveness of the deep learning and SSM approaches. There is no new data collected in this study and hence no human subjects or animals involved. The task is to localize the trabecular landmarks and the loss is the normalized sum of squared differences between test sample landmarks $\hat{\boldsymbol{s}}$ and the ground truth landmarks $\boldsymbol{s}^*$.

$$\begin{aligned} \text{loss} &= \frac{\mathcal{L}(s, s^*)}{\mathcal{N}(s^*)} \\ \mathcal{L}(s, s^*) &= \frac{1}{N_R} \sum_{i=1}^{N_R} \sqrt{(s_{i,x} - s^*_{i,x})^2 + (s_{i,y} - s^*_{i,y})^2} \end{aligned} \tag{5}$$

where $\mathcal{N}(s^*)$ is the normalization factor which equals to the average edge length of the ground truth bounding box. The smaller the loss is, the better the approach is.

A. Environmental Settings

We adopt 5-folds cross validation as the evaluation criteria and there are 80% training samples and 20% test samples in each fold. The final loss is calculated by averaging all of the 5 cross validation results.

B. Deep Learning Results

In deep learning approach, each sample is resized so that its edges are within $[800, 1333]$ pixels. The feature extractor is frozen and only the weights in detection head are finetuned. The network is trained for 60 epochs. The learning rate starts with 0.001 and is halved if the mean average precision (mAP) on validation set is not improved for three consecutive epochs. For fair comparisons, we propose two approaches, GD and LD, to convert detection results (*e.g.* bounding boxes) into landmark coordinates so that loss in Eq. (5) can be obtained. Dilation in LD approach is 2% of the image side length.

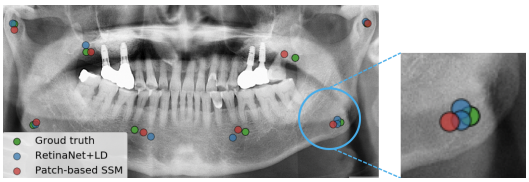


Fig. 3: Prediction result of each method.

Both the results of GD and LD on validation sets are reported. Table I enumerates the loss of 5-folds cross validation. LD outperform GD for most cases and the mean loss of LD surpasses GD's by a large margin. This could be due to the fact that trabecula landmarks lie within a certain range of DPR image that can be estimated approximately by the landmark distribution in training set. And thus, more potential landmark candidates can be selected rather than outliers lying far from the ground truth landmark.

C. SSM Results

Considering that the sizes of all DPR images are similar, we construct SSM model with only one scale. All images are firstly resized so that the length of diagonal is 300 pixels and then the SIFT feature [16] is extracted to construct appearance model. As shown in Figure 2, the center rectangle that 10% apart from each side is the initial bounding box. In this experiment, we apply both the holistic and patch-based SSM [25], [2] for locating landmarks. The holistic SSM uses a holistic appearance representation while the patch-based model extracts features at each landmark and then concatenates these features as an appearance instance.

The patch-based model ignores background noises and is thus more robust in case of limited dataset. Table I also enumerates the results of the holistic and patch-based SSM. We can see that patch-bases SSM outperforms all the other methods by a significant margin and achieves the lowest loss. This demonstrates the generalization capability and effectiveness of patch-based SSM.

The holistic and patch-based SSM locate landmarks better than RetinaNet with GD. Though RetinaNet with LD and the holistic SSM achieve similar loss, the latter does not require any post-processing and performs more stable. Besides, in experiments, we observe that the displacement of SSM predictions is much smaller than RetinaNet, as illustrated in Figure 3, which is partially because SSM takes use of spacial relationship between landmarks.

IV. CONCLUSIONS

In this paper, we develop an automatic trabecular landmark detection algorithm in the context of DPR-based osteoporosis prescreening. In order to compare our SSM with popular CNN-based detectors, we also propose two methods, GD and LD, for converting detection boxes into landmarks. In experiment on 108 DPR images, our solution is more robust and

TABLE I: Loss of 5-folds Cross Validation (in pixel)

Methods	CV-1	CV-2	CV-3	CV-4	CV-5	Median	Mean
RetinaNet+GD	0.0505	0.1095	0.0550	0.0554	0.0521	0.0550	0.0645
RetinaNet+LD	0.0395	0.0495	0.0631	0.0379	**0.0388**	0.0395	0.0458
Holistic SSM	0.0484	0.0430	0.0419	0.0496	0.0457	0.0457	0.0457
Patch-based SSM	**0.0384**	**0.0400**	**0.0342**	**0.0375**	0.0407	**0.0384**	**0.03816**

shows outstanding detection accuracy in comparison with state-of-the-art detection algorithms. Furthermore, compared with CNN-based detectors, our method avoids the need of a large amount of training data, which is more practical in application. The promising result encourages future work for automatic inexpensive osteoporosis prescreening.

REFERENCES

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikainen. Face description with local binary patterns: Application to face recognition. *TPAMI*, 28(12):2037–2041, 2006.

[2] Joan Alabort-i-Medina, Epameinondas Antonakos, James Booth, Patrick Snape, and Stefanos Zafeiriou. Menpo: A comprehensive platform for parametric image alignment and visual deformable models. In *ACM-MM*, 2014.

[3] Simon Baker and Iain Matthews. Lucas-kanade 20 years on: A unifying framework. *IJCV*, 56(3):221–255, 2004.

[4] Chunjuan Bo, Xin Liang, Peng Chu, Jonathan Xu, Dong Wang, Jie Yang, Vasileios Megalooikonomou, and Haibin Ling. Osteoporosis prescreening using dental panoramic radiographs feature analysis. In *ISBI*, 2017.

[5] Zhaowei Cai and Nuno Vasconcelos. Cascade r-cnn: Delving into high quality object detection. In *CVPR*, 2018.

[6] Peng Chu, Chunjuan Bo, Xin Liang, Jie Yang, Vasileios Megalooikonomou, Fan Yang, Bingyao Huang, Xinyi Li, and Haibin Ling. Using octuplet siamese network for osteoporosis analysis on dental panoramic radiographs. In *EMBC*, 2018.

[7] Timothy F Cootes, Gareth J Edwards, and Christopher J Taylor. Active appearance models. *TPAMI*, 23(6):681–685, 2001.

[8] Ross Girshick. Fast r-cnn. In *ICCV*, 2015.

[9] Keith Horner and Hugh Devlin. The relationship between mandibular bone mineral density and panoramic radiographic measurements. *Journal of dentistry*, 26(4):337–343, 1998.

[10] Muthu Subash Kavitha, Akira Asano, Akira Taguchi, Takio Kurita, and Mitsuhiro Sanada. Diagnosis of osteoporosis from dental panoramic radiographs using the support vector machine method in a computer-aided system. *BMC Medical Imaging*, 12(1):1, 2012.

[11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[12] André Ferreira Leite, Paulo Tadeu de Souza Figueiredo, Cláudio Mares Guia, Nilce Santos Melo, and Ana Patrícia de Paula. Correlations between seven panoramic radiomorphometric indices and bone mineral density in postmenopausal women. *Oral Surgery, Oral Medicine, Oral Pathology, Oral Radiology, and Endodontology*, 109(3):449–456, 2010.

[13] Peiyi Li, Xiong Yang, Fangfang Xie, Jie Yang, Erkang Cheng, Vasileios Megalooikonomou, Yong Xu, and Haibin Ling. Trabecular texture analysis in dental cbct by multi-roi multi-feature fusion. In *ISBI*, 2014.

[14] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *ICCV*, 2017.

[15] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In *ECCV*, 2016.

[16] David G. Lowe. Object recognition from local scale-invariant features. In *ICCV*, 1999.

[17] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *IJCAI*, 1981.

[18] Marcel Lthi, Andreas Forster, Thomas Gerig, and Thomas Vetter. Shape modeling using gaussian process morphable models. In Guoyan Zheng, Shuo Li, and Gabor Szkely, editors, *Statistical Shape and Deformation Analysis*, chapter 7, pages 165–191. Academic Press, 2017.

[19] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *CVPR*, 2016.

[20] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.

[21] Martin G Roberts, James Graham, and Hugh Devlin. Image texture in dental panoramic radiographs as a potential biomarker of osteoporosis. *IEEE Trans. Biomed. Engineering*, 60(9):2384–2392, 2013.

[22] Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. *ICLR*, 2014.

[23] Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. Subspace learning from image gradient orientations. *TPAMI*, 34(12):2454–2466, 2012.

[24] Konstantinos Z Vlasiadis, Chris A Skouteris, George A Velegrakis, Ivoni Fragouli, John M Neratzoulakis, John Damilakis, and Eugenios E Koumantakis. Mandibular radiomorphometric measurements as indicators of possible osteoporosis in postmenopausal women. *Maturitas*, 58(3):226–235, 2007.

[25] Matthias Wilms, Heinz Handels, and Jan Ehrhardt. Representative patch-based active appearance models generated from small training populations. In *MICCAI*, 2017.

[26] Nicole C. Wright, Anne C. Looker, Kenneth G. Saag, Jeffrey R. Curtis, Elizabeth S. Delzell, Susan Randall, and Bess Dawson-Hughes. The recent prevalence of osteoporosis and low bone mass in the united states based on bone mineral density at the femoral neck or lumbar spine. *Journal of Bone and Mineral Research*, 29(11):2520–2526, 2014.

[27] Sijia Yu, Peng Chu, Jie Yang, Bingyao Huang, Fan Yang, Vasileios Megalooikonomou, and Haibin Ling. Multitask osteoporosis prescreening using dental panoramic radiographs with feature learning. *Smart Health*, in press.