

Fully automatic teeth segmentation in adult OPG images^{* **}

N. Vila-Blanco^{1[0000-0001-5865-9973]}, T. Cootes^{2[0000-0002-2695-9063]}, C. Lindner^{2[0000-0001-9396-3436]}, I. Tomás^{3[0000-0002-3317-0853]}, and M.J. Carreira^{1[0000-0003-0532-2351]}

¹ Centro de Investigación en Tecnologías da Información (CITIUS), Universidade de Santiago de Compostela, España.

² Centre for Imaging Sciences, University of Manchester, U.K.

³ Oral Sciences Research Group, Universidade de Santiago de Compostela, Health Research Institute Foundation of Santiago (FIDIS), Spain.

Abstract. This work addresses the problem of segmenting teeth in panoramic dental images. Random Forest Regression Voting Constrained Local Models (RFRV-CLMs) were applied firstly to locate the mandible and the approximate pose of each tooth, and secondly to locate the full outline of each individual tooth. An automatically computed quality-of-fit measure was proposed to identify missing teeth. The system was evaluated using 346 manually annotated images containing adult-stage mandibular teeth. Encouraging results were achieved for detecting missing teeth. The system achieved state-of-the-art performance in locating the outline of present teeth with a median point-to-curve error of 0.2mm for each of the teeth.

Keywords: Teeth segmentation · panoramic dental images · random forest regression-voting · machine learning.

1 Introduction

Dental radiographs have been widely used since the discovery of X-rays in a variety of fields: abnormality detection, treatment and/or surgery planning, prostheses design, assessment of children's dental development, human identification by dental matching, and many more. X-ray images provide additional information to the simple exploration of the oral cavity since they reveal hidden parts of the teeth and other surrounding structures. There are several types of dental x-ray images depending on the captured area. In intraoral images, the sensor is placed inside the mouth and the images cover some specific area (no more than 3-4

^{*} This work has received financial support from the Consellería de Cultura, Educación e Ordenación Universitaria (accreditation 2016-2019, ED431G/08, growth potential group 2017-2020 ED431B 2017/029, reference competitive group 2017-2020, ED431C 2017/69, and N Vila-Blanco support ED481A-2017) and the European Regional Development Fund (ERDF).

^{**} C. Lindner is funded by the Medical Research Council, UK (MR/S00405X/1).

complete teeth). In contrast, in extraoral images the sensor is placed outside the mouth and the images cover a bigger area. That is the case for panoramic images, which provide a complete coverage of the dentition and other surrounding bones and tissues with a very small dose of ionising radiation. Although their quality is highly dependent on patient positioning and patient movements during acquisition [1, 2] they have been widely used to diagnose periodontal disease, cysts in the jaw bones, jaw tumours, oral cancer, impacted teeth, temporomandibular joint disorders or sinusitis, among others.

One of the key tasks in automatic dental image processing is teeth segmentation. This has proven to be useful in a variety of areas such as human identification [3–5], caries detection [6], lesion detection [7] or even dental age estimation [8]. The works in this area tackled automatic or semiautomatic teeth segmentation mostly from intraoral images in a variety of ways, comprising thresholding [4, 9], combination of morphological operations [10, 11], active contours [3], level sets [12], Mixture of Gaussians [5] and many more. Although these algorithms can reach great performance in a variety of applications, they present some problems when working with dental images, mainly because they are very sensitive to intensity changes, dental restorations, teeth injuries and overlapping teeth. Thus, there is a need to follow more robust approaches which use domain knowledge to improve the results.

In this regard, methods utilising statistical models have proven to be accurate and robust in medical image segmentation. One of latest contributions on this area are Random Forest Regression-Voting Constrained Local Models (RFRV-CLMs) [13], which are an evolution of the original Constrained Local Models [14] and combines a global shape model with individual point appearance models. Over the last years, this approach has been applied to a variety of medical images with high performance [15–17], which encourages us to use it in the teeth segmentation problem.

Our main contribution is the development of a fully automatic procedure to outline mandibular adult-stage teeth in panoramic images, including the identification of any missing teeth.

2 Methods

2.1 RFRV-CLM

Random Forest Regression-Voting Constrained Local Models combine a linear shape model with a set of local models designed to locate each point. RFRV-CLMs are summarised in the following, and the reader is referred to [13, 15]) for full details.

Each annotated shape is encoded as a vector $x = (x_0, y_0, x_1, y_1, \dots, x_{n-1}, y_{n-1})^T$ with the concatenated coordinates of the n landmark points. In order to train the model, the shapes are resampled and aligned in a reference frame so a linear model can be built as follows:

$$x = T_\theta(\bar{x} + Pb_i + r_i) \quad (1)$$

where \bar{x} is the mean shape, P are eigenvectors of the covariance matrix, b is a vector of shape parameters, r is a regularisation term which allows small deviations from the model and T_θ is a similarity transformation of parameters θ which maps the shape from the reference frame to the image frame.

In order to locate each individual point, Random Forest Regression-Voting is used. The region of interest which encloses all landmark points is resampled into a standardised reference frame and for each landmark point l in x a set of image patches p_j are sampled at random displacements d_j (i.e. centred at $l + d_j$). Then a set of decision tree regressors are trained from the Haar features [18] of all patches to predict the displacements.

Given a new image and an initial estimation of the pose of the mean shape, the region of interest is resampled into a standardised reference frame and a set of image patches are sampled at random displacements around each initial estimated point. Haar features are extracted from the patches and fed into the RF regressors. The outputs of all decision trees are accumulated in a voting grid V_l , where the positions of the grid with higher values indicate the most likely position for that landmark point.

The local appearance models and the global shape model are combined as follows:

$$\begin{aligned} Q(b, \theta, r) &= \sum_{l=0}^{n-1} V_l(T_\theta(\bar{x}_l + Pb_l + r_l)) \\ \text{s.t. } b^T S_b^{-1} b &\leq M_t \quad \text{and} \quad |r_l| < r_t \end{aligned} \quad (2)$$

where M_t and r_t are thresholds on the Mahalanobis distance and the regularisation term, respectively, and S_b is the covariance matrix of the shape model parameters b . This yields the overall quality-of-fit (QoF) measurement Q (Eq. 2), which represents the total number of votes for a shape defined by parameters $\{b, \theta, r\}$.

The search process is carried out iteratively, so for each search iteration, the algorithm gets the set of parameters $\{b, \theta, r\}$ which maximises the overall QoF and updates the landmark points.

2.2 Two-step teeth segmentation

We build separate RFRV-CLMs for each tooth type. Given that the dentition is almost horizontally symmetric, a single model trained from one tooth on one side (left or right) can also be used to segment the corresponding tooth on the opposite side. It is worth mentioning that there are two main problems with teeth segmentation from individual teeth models. First of all, the space occupied by each tooth is very small when compared to the image size, so the search process requires a reasonably good initialisation. Furthermore, teeth of the same type (e.g. single-root and multi-root) are very similar to each other so the search process can easily end up converging to a neighbouring tooth.

To overcome these problems, in addition to individual teeth models, another model was trained from some keypoints in the image. The idea is to identify a set

of representative points in each tooth and the mandible which give a reasonably good approximation of their poses (see Table 1). Thus, this model is able to capture the pose variation of each tooth (in terms of position, size and rotation) in relation to neighbouring teeth and the mandible. As the mandible occupies a similar percentage in all panoramic images, a good initialisation of the search model can be carried out by placing the mean shape in the centre of the image and scaling it to the 75% of the image width.

The search process for a new image is performed fully automatically in two steps. In the first step, the keypoint model looks for the optimal localisation of the teeth and mandible keypoints. Then, the initial pose estimation of each tooth is carried out via (3):

$$\arg \min_{\theta} d(k_t, T_{\theta}(\bar{x}_k)) \quad (3)$$

where k_t is the estimation of the keypoints of tooth t provided by the first model, \bar{x}_k are the keypoints of the mean shape of tooth t and d is the Euclidean distance function. The initial shape estimation for each tooth is, therefore, the result of applying the estimated pose to the mean shape, $T_{\theta}(\bar{x})$.

On completion of the search, the quality-of-fit of each model point is calculated as the magnitude of the mean displacement vector produced by the random forest for the point when evaluated on a patch centred on the point. This should be small for good matches and larger for those points which don't match so well. To obtain a score for the whole tooth we compute the mean, m , and standard deviation, sd , of the values for each point, and construct the final score as $S = m + sd$. This has been shown to be a more effective discriminator than just using the mean alone. We treat a tooth as missing if S is below a threshold.

3 Experiments and results

In this work, a set of 346 panoramic images provided by the School of Medicine and Dentistry, Universidade de Santiago de Compostela (Spain) have been used, all of which were collected under ethical approval. To test the proposed segmentation approach, the images where one hemi-arch including all seven left-mandibular teeth (from the first incisor to the second molar) were present have been used as the train set, and the remaining images have been used as the test set. In total, 261 images have been used for training and 85 for testing. In each image, the shapes of seven left-mandibular teeth (from 31 to 37) have been manually annotated as well as 7 mandible keypoints (see Figure 1 and Table 1). In total, each training example consists of a set of 263 landmark points.

The individual tooth models and the keypoint model were built using the RFRV-CLM algorithm. The mean shape of each tooth model is shown in Figure 2. For each model, a coarse-fine approach has been followed, which in this case consists of training a fine model where the reference frame width is approximately the desired object width, and training a coarse model where the frame width is about a quarter of the fine frame width. This gives a rough but more robust shape estimation at first and then refines the shape. In the case of the keypoint

model, the search process consists of 3 search iterations with the coarse model and 2 search iterations with the fine model. For the individual teeth models, the iterations of coarse and fine searches have been reduced to 2 and 1, respectively.

Fig. 1: Annotated points in a single image. In red: teeth and mandible keypoints; in blue: teeth non-keypoints.

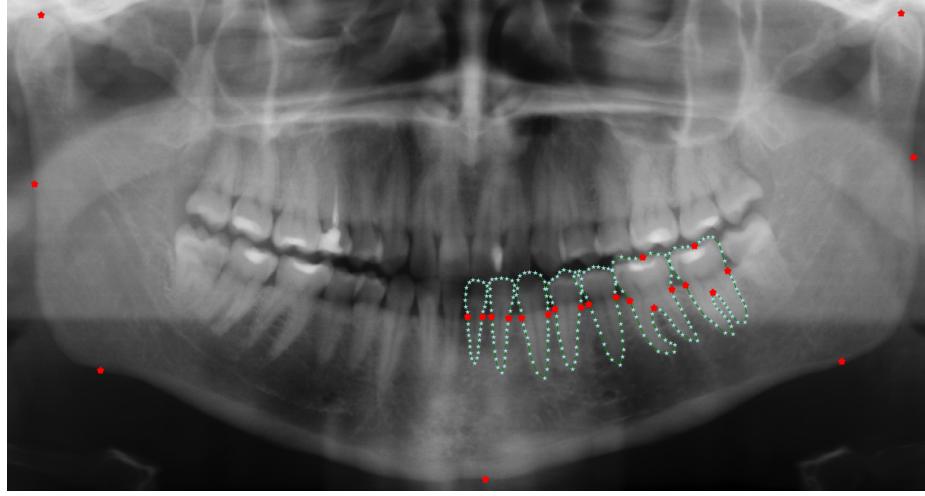
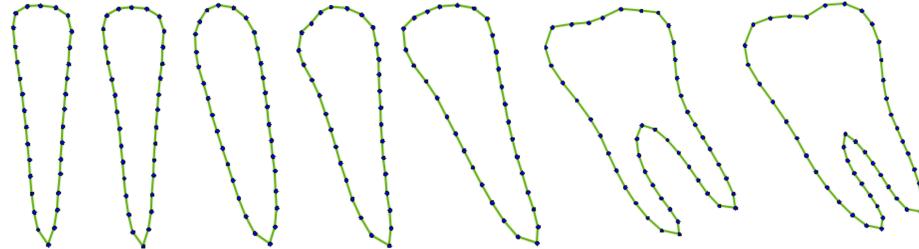


Table 1: Number of points used in each individual model and number of points of each model used in the keypoint model.

Model	#points	#points used in keypoint model
Central incisor (31)	32	2 (one on each neck side)
Lateral incisor (32)	32	2 (one on each neck side)
Canine (33)	32	2 (one on each neck side)
First premolar (34)	32	2 (one on each neck side)
Second premolar (35)	32	2 (one on each neck side)
First molar (36)	48	4 (one on each neck side, one at the top and one at the root bifurcation)
Second molar (37)	48	4 (one on each neck side, one at the top and one at the root bifurcation)
Mandible	7	7 (nearly equidistant points between the top of the condyles)

The predicted shapes of teeth 31 to 37 have been compared to manually annotated shapes and the performance of the proposed approach has been assessed in three ways. Firstly, the performance of present/missing teeth detection has been measured. Table 2 shows the classification results when choosing a threshold to maximise (True Positive Rate - False Positive Rate). See Figure 4 for some examples.

Fig. 2: Mean shape of each individual tooth model.



Secondly, to assess whether they have been located correctly, the Intersection over Union (IoU) of annotated and predicted shapes was calculated from the examples where both teeth are present and are correctly detected as present. Table 3 shows that the detection of multiroot teeth (36 and 37) is slightly more successful than the detection of single root teeth. This is likely to be because the anterior teeth are closer to each other so the model might match a neighbouring tooth. Assuming that an overlap greater than 50% between the prediction and the ground truth indicates that the predicted shape is very likely to match the real tooth, the examples with a IoU value over 0.5 have been treated as correctly located. In general, the proportion of well-located teeth is over 90% among all teeth types.

Table 2: Confusion matrix of the missing teeth detection problem and binary classification metrics. In order to obtain these metrics, the “present” class has been considered as the positive class.

		Prediction		Total	Accuracy	95.46%
Real		Missing	Present			
	Missing	31 (TN)	6 (FP)	37	Precision	98.9%
	Present	21 (FN)	537 (TP)	558	Sensitivity	96.24%
	Total	53	542	595	Specificity	83.78%

Thirdly, the accuracy of the tooth shape matching has been evaluated on the correctly located teeth (where the overlap between model and true tooth is greater than 50%) with the point-to-curve error, which represents the shortest distance from each estimated point to the curve through the ground truth landmark points (Table 4). The median of the errors is less than 0.23 millimetres for all types of teeth. The 99%ile is 1.31mm in the worst case, which demonstrates the robustness of the proposed segmentation approach.

Table 3: Intersection Over the Union statistics for each individual tooth predictions. In the last column, the percentage of the examples with an IoU over 0.5, which are treated as correctly located.

Model	mean	sd	median	% over 0.5
31	0.78	0.24	0.86	89.41%
32	0.80	0.20	0.86	92.94%
33	0.84	0.18	0.90	94.12%
34	0.84	0.22	0.91	91.76%
35	0.85	0.23	0.93	92.21%
36	0.87	0.18	0.92	95.77%
37	0.86	0.21	0.92	94.29%

Table 4: Point to curve statistics in each individual tooth model (in mm): mean, standard deviation, median and 90, 95 and 99 centiles. These results have been obtained on the examples where the teeth have been correctly located.

Model	mean	sd	med	90%	95%	99%
31	0.23	0.12	0.21	0.38	0.47	0.61
32	0.30	0.23	0.22	0.58	0.65	1.29
33	0.30	0.24	0.23	0.56	0.76	1.31
34	0.24	0.20	0.19	0.36	0.56	0.90
35	0.21	0.11	0.18	0.30	0.48	0.58
36	0.25	0.16	0.20	0.35	0.50	0.89
37	0.27	0.14	0.23	0.41	0.48	0.82

Note that all performance measurements have been obtained on the left mandibular teeth only as we did not have manual ground truth annotations for the right side. However, the right mandibular teeth can be outlined by applying the left mandibular teeth models to the horizontally reflected images. See Figure 3 for some examples.

4 Discussion and conclusions

We have shown that a state-of-the-art performance can be achieved in adult mandibular teeth segmentation by using the RFRV-CLM algorithm in two steps. The first step provides an estimation of some teeth and mandible keypoints, which are used to initialise each individual tooth search. In the second step, the search of each tooth is performed independently. This two-step approach overcomes the problem of automatically initialising each individual tooth model, and the results show that the teeth shapes can be matched very accurately, especially if the tooth is correctly located.

A limitation of this study is that we have not taken into account the third molar (also known as the *wisdom tooth*). This is because this tooth is often extracted or missing in some patients so we had very few examples. Moreover,

although the QoF statistics are a good starting point for missing teeth detection, this task could be improved by using other metrics or algorithms developed specifically for that purpose.

Nonetheless, the presented results are promising and are a big step towards a fully automatic dental assessment tool with a variety of applications. Two direct uses of the proposed system are (i) automatic teeth measurements with a view to planning surgical treatments; and (ii) automatic radiograph matching with the aim of identifying people (e.g. in forensics). Other clinical tasks could also be carried out with this system and few functionality additions. For example, the detection of caries, impacted tooth and other abnormalities.

References

1. R. H. N. Rondon, Y. C. L. Pereira, and G. C. do Nascimento, “Common positioning errors in panoramic radiography: A review,” *Imaging Science in Dentistry*, vol. 44, no. 1, pp. 1–6, 2014.
2. M. Halperin-Sternfeld, E. E. Machtei, C. Balkow, and J. Horwitz, “Patient movement during extraoral radiographic scanning,” *Oral Radiology*, vol. 32, no. 1, pp. 40–47, 2016.
3. H. Chen and A. K. Jain, “Tooth contour extraction for matching dental radiographs,” in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, vol. 3, pp. 522–525, IEEE, 2004.
4. O. Nomir and M. Abdel-Mottaleb, “A system for human identification from x-ray dental radiographs,” *Pattern Recognition*, vol. 38, no. 8, pp. 1295–1305, 2005.
5. H. Chen and A. K. Jain, “Dental biometrics: Alignment and matching of dental radiographs,” in *Application of Computer Vision, 2005. WACV/MOTIONS’05 Volume 1. Seventh IEEE Workshops on*, vol. 1, pp. 316–321, IEEE, 2005.
6. J. Oliveira and H. Proen  a, “Caries detection in panoramic dental x-ray images,” in *Computational Vision and Medical Image Processing*, pp. 175–190, Springer, 2011.
7. S. Li, T. Fevens, A. Krzy  ak, C. Jin, and S. Li, “Semi-automatic computer aided lesion detection in dental x-rays using variational level set,” *Pattern Recognition*, vol. 40, no. 10, pp. 2861–2873, 2007.
8. L.   ular, M. Tomai  , M. Suba  i  , T.   ari  , V. Sajkovi  , and M. Vodanovi  , “Dental age estimation from panoramic x-ray images using statistical models,” in *Image and Signal Processing and Analysis (ISPA), 2017 10th International Symposium on*, pp. 25–30, IEEE, 2017.
9. N. S. Ahmad, Z. M. Zaki, W. Ismail, et al., “Region of adaptive threshold segmentation between mean, median and otsu threshold for dental age assessment,” in *Computer, Communications, and Control Technology (I4CT), 2014 International Conference on*, pp. 353–356, IEEE, 2014.
10. P. H. Lira, G. A. Giraldi, and L. A. Neves, “Panoramic dental x-ray image segmentation and feature extraction,” in *Proceedings of V Workshop of Computing Vision, Sao Paulo, Brazil*, 2009.
11. Y. Y. Amer and M. J. Aqel, “An efficient segmentation algorithm for panoramic dental images,” *Procedia Computer Science*, vol. 65, pp. 718–725, 2015.
12. S. Shah, A. Abaza, A. Ross, and H. Ammar, “Automatic tooth segmentation using active contour without edges,” in *2006 Biometrics Symposium: Special Session on Research at the Biometric Consortium Conference*, pp. 1–6, IEEE, 2006.

13. C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, “Robust and accurate shape model matching using random forest regression-voting,” *IEEE transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1862–1874, 2015.
14. D. Cristinacce and T. Cootes, “Automatic feature localisation with constrained local models,” *Pattern Recognition*, vol. 41, no. 10, pp. 3054–3067, 2008.
15. C. Lindner, S. Thiagarajah, J. Wilkinson, T. Consortium, G. Wallis, and T. Cootes, “Fully automatic segmentation of the proximal femur using random forest regression voting,” *IEEE Transactions on Medical Imaging*, vol. 32, no. 8, pp. 1462–1472, 2013.
16. T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer, “Robust and accurate shape model fitting using random forest regression voting,” in *European Conference on Computer Vision*, pp. 278–291, Springer, 2012.
17. P. Bromiley, J. Adams, and T. Cootes, “Localisation of vertebrae on dxa images using constrained local models with random forest regression voting,” in *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, pp. 159–171, Springer, 2015.
18. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 1, pp. I–I, IEEE, 2001.

Fig. 3: Results of automatic teeth segmentation process. In red, the predicted shapes. In blue, the left-mandibular manually annotated shapes. The segmentation is robust to some issues such as very bright images (b) or tooth filling (c). It also can manage teeth overlapping. It is worth noting that the most noticeable segmentation errors are observed in the apical regions (around root apices) due to the low contrast of the image in that area.

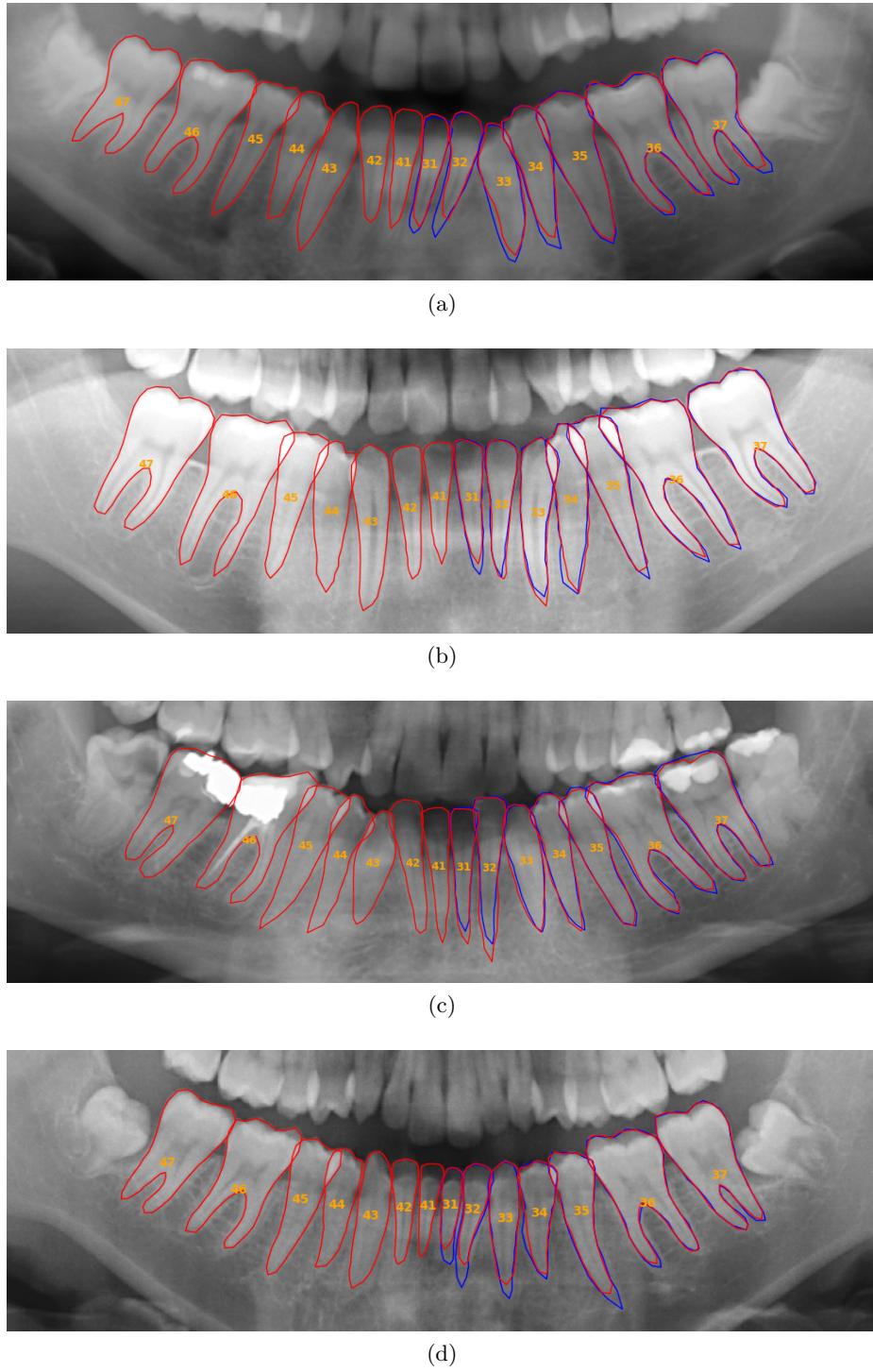
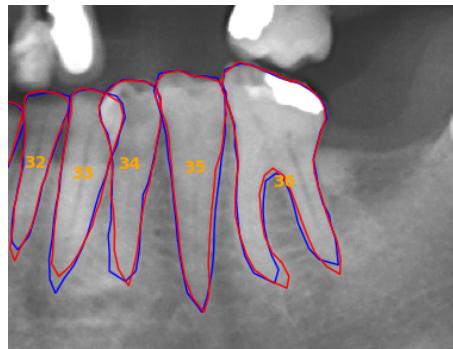


Fig. 4: Results of present/missing teeth detection. In red, the predicted shapes. In blue, the left-mandibular manually annotated shapes.



(a) Missing tooth 36 detected correctly.



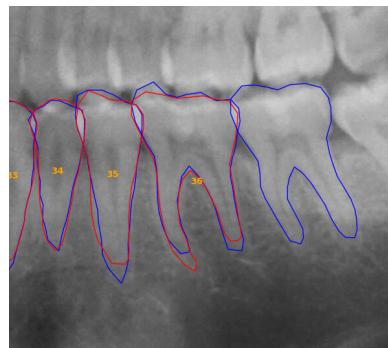
(b) Missing tooth 37 detected correctly.



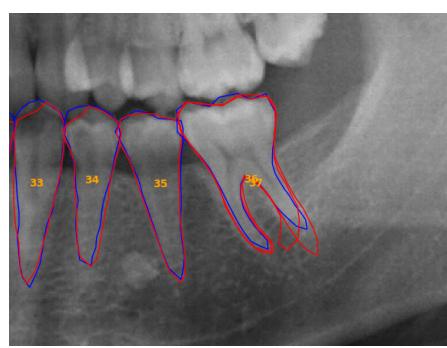
(c) Missing tooth 35 detected correctly.



(d) Tooth 35 searcher detects correctly tooth as missing. This is a special clinic case called agenesis, where the tooth (premolar 35 in the image) fails to form, so the related primary tooth is not pushed outward.



(e) Tooth 37 detected incorrectly as missing.



(f) Tooth 37 detected incorrectly as present due to it matches tooth 36.