Alveolar Bone Detection from Dental Cone Beam Computed Tomography using YOLOv3-tiny

Monica Widiasri
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
monicawidiasri.207025@mhs.its.ac.id
Department of Informatics
University of Surabaya
Surabaya, Indonesia
monica@staff.ubaya.ac.id

Agus Zainal Arifin
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
agusza@if.its.ac.id

Nanik Suciati
Department of Informatics
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
nanik@if.its.ac.id

Eha Renwi Astuti
Department of Dentomaxillofacial
Radiology
Universitas Airlangga
Surabaya, Indonesia
eharenwi@gmail.com

Rarasmaya Indraswari
Department of Information Systems
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia
raras@if.its.ac.id

Abstract—**Cone Beam Computed Tomography (CBCT) is a medical imaging technique widely used in dentistry including dental implant planning. To determine the size of the dental implant, it is necessary to detect the alveolar bone at the implant site. In this study, we propose automatic detection of alveolar bone from CBCT images of teeth using the YOLOv3-tiny method. The YOLOv3-tiny network architecture consists of a seven-layer convolution networks and six max-pooling layers in the Darknet-53 network with two output branch scale predictions. CBCT images of teeth obtained from 4 patients consisted of 800 coronal slices of 2D grayscale images, containing 830 alveolar bone annotations. Before the training process, the ground truth image annotation was made in the form of a bounding box on the alveolar bone object. The detection results of the YOLOv3-tiny model were compared with the detection results of the YOLOv3 and YOLOv2-tiny models. The results of the experiment on 640 training images and 160 testing images showed that YOLOv3-tiny outperformed YOLOv2-tiny with mAP of 98.6% and 96.73%, respectively. Meanwhile, shows the same good result as YOLOv3.**

Keywords— dental implants, CBCT, alveolar bone, YOLOv3-tiny, object detection

I. Introduction

Atrophy of the alveolar bone occurs after the loss of teeth. Dental implants have become a reliable treatment option for the replacement of missing teeth and preventing bone atrophy [1]. The success of dental implants depends on the quality and quantity of bone available at the recipient site, namely the height and width [2][3]. Cone-Beam Computed Tomography (CBCT) is often used in dental implant planning, as it provides excellent detail visualization of alveolar bone with clear boundary definitions [3][4]. Using CBCT images, the radiologist performs manual measurements of the alveolar bone height and width in each cross-sectional image using 3D imaging software. However, measuring alveolar bone from a large number of slices in CBCT is a time-consuming task and the accuracy is operator-dependent. Therefore, it is necessary to have an automatic detection system in dental implant planning to reduce the workload of radiologists and to detect alveolar bone accurately.

Recently, deep learning has been used to enhance the segmentation and detection performance of dental images.

Segmentation of alveolar bone and finding alveolar crests using U-net convolutional neural network can help automate alveolar bone segmentation from ultrasound images [5][6]. The proposed method yielded better performance than a state-of-the-art graph cuts-based method. An automated tooth segmentation method based on the U-Net structure in CBCT for dental implant planning improved the segmentation performance [7]. A novel deep learning hybrid method was proposed to automatically detect and classify the periodontal bone loss of each individual tooth on dental panoramic radiographs [8]. However, the application of deep learning to detect alveolar bone from CBCT images remains to be explored.

A state-of-the-art object detection system, You Only Look Once (YOLO), uses deep learning to detect and classify objects at the same time [9]. YOLOv3 is the latest version of the YOLO method proposed by Redmon, et al [10]. The YOLOv3 is used to detect masses that may be present in breast mammograms, successfully detected 89.4% of masses on INbreast mammograms with an average precision of 94.2% for benign masses and 84.6% for malignant masses [10]. YOLOv3-tiny was developed to simplify the network structure and reduce the YOLOv3 parameter [10]. To determine the dimensions of dental implants, automatic detection of alveolar bone must be carried out accurately in the planning of dental implants from CBCT images.

In this research, we propose an automatic detection of the alveolar bone using the YOLOv3-tiny detector model for CBCT images. The YOLOv3-tiny model was chosen mainly because it is fast and yields more accurate results than earlier versions. The automation of the detection process of the alveolar bone makes it easy to calculate the height and width of the alveolar bone in the dental implant area. This can make dental implant planning easier and more efficient and can reduce the workload of radiologists.

The rest of this paper is structured as follows. Section II presents the CBCT dataset used for the experiments and methods used in this study. Section III describes the experiments that have been performed and are followed by a discussion of the experiment's results. The conclusion and future work of this study is given in Section IV.

II. MATERIAL AND METHOD

We used dental Cone Beam Computed Tomography images of the mandible in the coronal view for input images. The YOLOv3-tiny method is used to detect alveolar bones.

A. Dataset

Cone Beam Computed Tomography (CBCT) is a medical imaging technique that produces multi slices of the 2D image that can be reconstructed into a 3D image. CBCT is commonly used in dental and radiological practices due to some limitations of conventional CT [2][11]. CBCT can be visualized using multiplanar reformation as axial (maxilla to mandibular), coronal (anterior to posterior), and sagittal (left to right) view [4]. CBCT imaging is used to assist in the planning and processing of dental implants. CBCT helps the surgeon to accurately measure the available bone in the dental implant planning. Bone quantity and quality in the implant placement site influence the accuracy of implant placement [12].

In this study, we focused on the posterior site. From CBCT images of the posterior mandible on the coronal view, we can measure the height and width of the alveolar bone to determine the size of the implant. Alveolar bones are part of the jaw bone that hold the teeth [5]. An example of an alveolar bone object marked with a box surrounding it on the CBCT image of the coronal slice can be seen in Fig. 1. The research used dental CBCT images of 4 patients who were planning for implants from the Dental and Oral Hospital of Airlangga University (RSGM UNAIR). CBCT images were obtained with Instrumentarium OP-300 3D X-ray unit The 3D images had the size of $266 \times 266 \times 200$ voxels. For each patient, there are 200 coronal slices of 2D grayscale images. Thus, the total data used for this study were 800 images. Divide the images into training sets and test sets, with a ratio of 80%:20%. The training process used 640 images, while 160 images will be used for testing.

B. Ground Truth Preparation

Ground truth images are created by labeling objects with a bounding box made using the LabelImg tool, a graphic annotation tool. The annotated ground truth was confirmed by a dental radiologist expert at RSGM UNAIR. The annotation
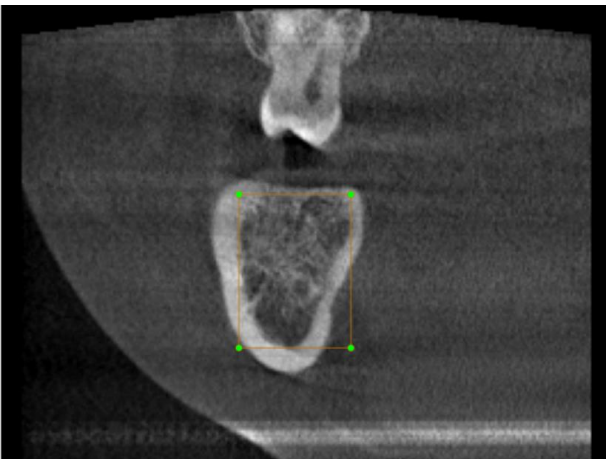

Fig. 1. Example of an alveolar bone annotation from CBCT image on coronal slices

process generates a text file for each image, containing the class number and coordinate values relative to the image width and height for each object in it. Format annotation is (object-id) (x_c) (y_c) (width) (height) in each line for each object [13]. Fig. 1 shows an example of annotated ground truth image. The alveolar bone annotation is represented by a yellow rectangle with four green vertices. The text file of the annotated image in Fig. 1 consists of

(0) (0.483) (0.583) (0.184) (0.335).

The meaning of that text annotation is object-id=0 represents an object alveolar bone, coordinates center of the bounding box relative to the width and the height of image are x_c=0.483 and y_c=0.583, the width of bounding box relative to the width of image is 0.184, and the height bounding box relative to the height of image is 0.335.

Preparation of ground truths resulted in 830 annotations of the alveolar bone. A total of 663 annotations for the training dataset and 167 annotations for the testing dataset.

C. Detection Process

The YOLO network is a one-stage object detection algorithm, proposed by Redmon, et al [9]. YOLO uses a single neural network to process the full image for both classification and localization. YOLO is one of the fastest object detection methods with high accuracy and good performance [14]. YOLOv3, the last version of the YOLO method proposed by Redmon, et al, uses Darknet-53 architecture, as the feature extractor, feature pyramid network, and binary cross-entropy loss to improve the detection accuracy [15].

We detect alveolar bone with YOLOv3-tiny. The YOLOv3-tiny is a compressed YOLOv3 network model for constrained environments and faster training. It is composed of seven-layer convolution networks and six max-pooling layers in the Darknet-53 network [14]. It also reduces the output branch from the three-scale predictions to two-scale predictions. At each scale, three anchor boxes are used to predict three bounding boxes for any grid cell. The YOLOv3-tiny network structure can be seen in Fig. 2. In this study, the yolov3-tiny.conv.11 is loaded in YOLOv3-tiny before training as pretrained weights.

The detection process of YOLOv3-tiny is as follows, first, resize an input image to 416 pixels × 416 pixels. Second, divide the input image into 13×13 grids. Then, every grid will use three bounding boxes to detect an object. It will generate $13 \times 13 \times 3$ bounding boxes for an input image. If the center of several objects is on a grid, then the network outputs a confidence score that describes the confidence level of the grid containing objects and the accuracy of the bounding box when it contains objects [16]. Also outputs offset values for the bounding box on the grid in predicting objects. The bounding box which has a confidence score above the threshold value is selected and used to locate objects in the image [14].

III. RESULT AND DISCUSSION

The mAP (mean value of average precision), precision, recall, and F1-score are used to quantitatively evaluate the performance of different methods. The calculations of the mAP, precision, recall, and F1-score, were shown in the following Eqs (1)-(4). Here, the mAP is the mean of the AP (Average Precision) for all the classes, when alveolar bone or
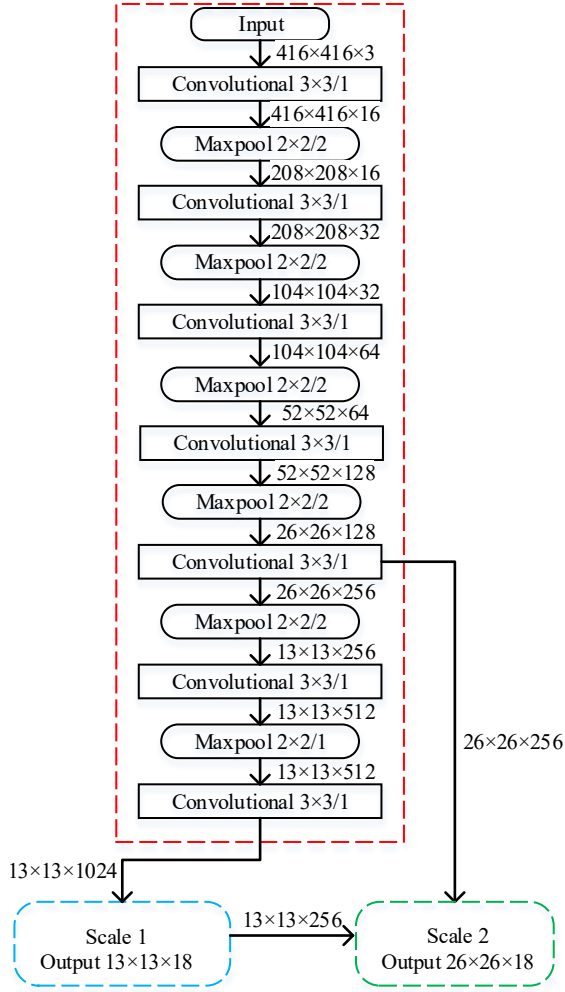
Fig. 2. YOLOv3-tiny network structure

mandibular canal is detected; and the higher the value is, the better the detection result.

$$mAP = \frac{\sum_{c=1}^{C} AP(c)}{C} \qquad (1)$$

$$Precision = \frac{TP}{TP+FP} \qquad (2)$$

$$Recall = \frac{TP}{TP+FN} \qquad (3)$$

$$F1 - score = 2x \frac{Precision \; x \; Recall}{Precision + Recall} \qquad (4)$$

where TP, FP, and FN are the numbers of true positive cases, false positive cases, and false negative cases, respectively; and C is the number of detection classes. In this research C=1, there is one class (i.e. alveolar bone).

A. YOLOv3-tiny Performance Training

The training experiments for the alveolar bone detection model using YOLOv3-tiny have been carried out up to 4000 iterations (max batch = 4000). The training process curve can be seen in Fig. 3. The red line shows the change in the value of losses during training. At the beginning of the training,

Metode	TP	FP	FN	Prec.	Recall	F1-score	mAP (%)	Train. time (hours)
YOLOv2-tiny	159	5	8	0.97	0.95	0.96	96.73	0.94
YOLOv3	164	0	3	1	0.98	0.99	98.6	6.1
YOLOv3-tiny	**162**	**4**	**5**	**0.98**	**0.97**	**0.97**	**98.6**	**0.85**

approximately up to 200 iterations, the loss value is still high, meaning that the model has not been able to detect the alveolar bone object correctly. But after that until the 400th iteration, the value of the loss was drastically reduced. This shows that during the iteration the model can begin to detect the alveolar bone object correctly. The value of the loss gradually decreases up to 1200 iterations. After that the value of the loss fluctuates slightly around 0.2.

On this curve, the red line represents the change in the mAP value. The mAP value is displayed from the 1000th iteration which is 82%. After the 1000th iteration, the mAP value increased significantly, and around the 1100th iteration, the mAP is above 90%. And when it reaches the 2000th iteration, the mAP value fluctuates slightly around 99%.

These results indicate that with a maximum of 2000 iterations can achieve effective detection results. The loss and mAP values until the iteration reach a good accuracy value of around 99% with a loss value of around 0.2. This is in accordance with what was suggested in [13], the maximum number of batches used was 2000 * the number of classes, where in this study the number of classes used was 1, so the experiment in this study used a max batch of 2000.

B. Performance Comparison with other YOLO method

To further evaluate the performance of YOLOv3-tiny, the detection performance of alveolar bone for the test data was compared with the detection system using the other YOLO method, namely YOLOv3-tiny and YOLOv2-tiny. The training process is carried out up to a maximum batch of 2000. The training result detector model will be used to detect the test data. The comparison results of TP, FP, FN, Precision, Recall, F1-score, and mAP are shown in Table I.

From the results of the experiment, the mAP value of YOLOv3-tiny was as good as YOLOv3 and better than the YOLOv2-tiny, respectively 98.6% and 96.73%. As for the precision, recall, and F1-score values, the detection model using YOLOv3 outperformed other methods. However, when compared to tiny YOLO, alveolar bone detection using YOLOv3-tiny is better than YOLOv2-tiny. YOLOv2-tiny is designed to reduce the Darknet19 network from YOLOv2 to nine layers and six unified layers at one scale. While the architecture for feature extraction on YOLOv3-tiny which uses two scales is greater than YOLOv2-tiny, this method allows YOLOv3-tiny to obtain more meaningful semantic information, thereby increasing detection accuracy.

The comparison of the training time required to build an alveolar bone detector model can also be seen in Table I. From these results, it can be seen that the training time to build the
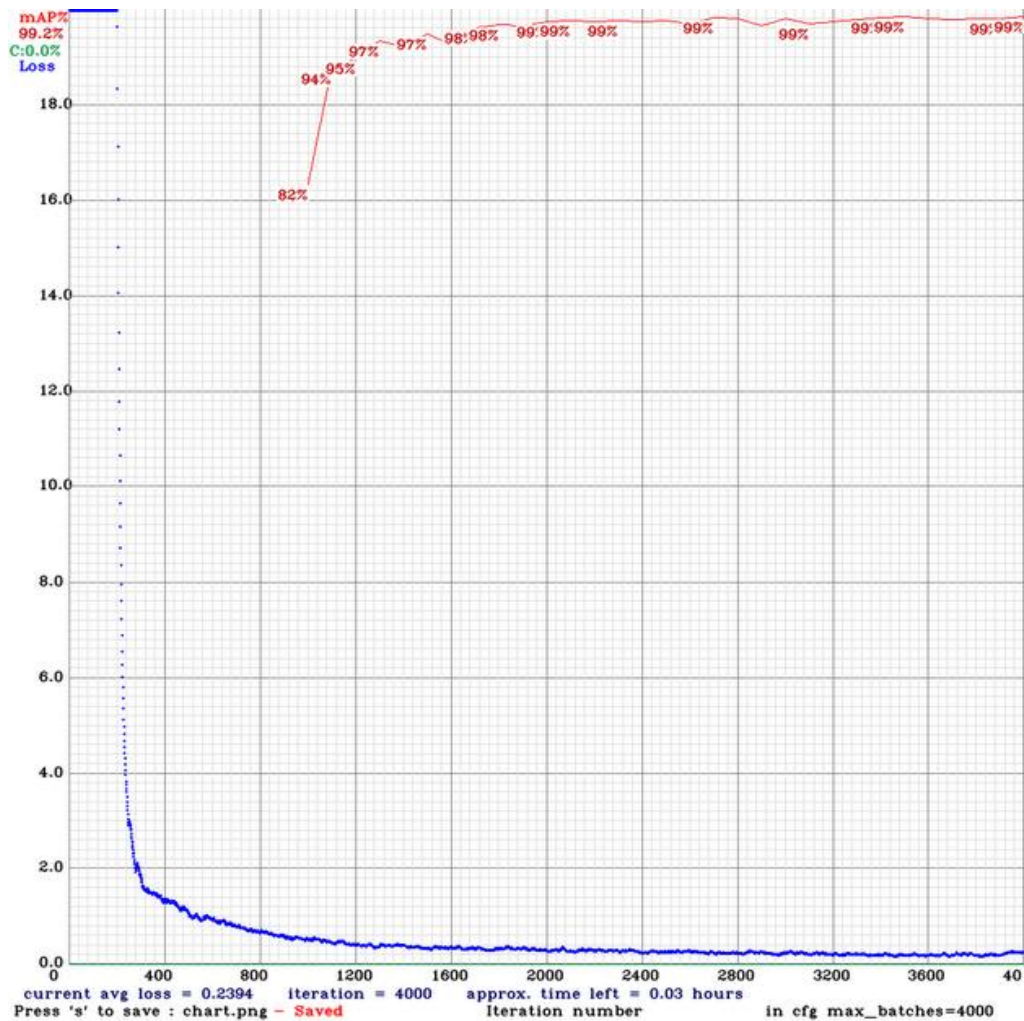
Fig. 3.   YOLOv3-tiny performance training

YOLOv3-tiny detector model is faster than the YOLOv2-tiny detector model (both of which are tiny YOLO detector models). Darknet-53 used in YOLOv3 and YOLOv3-tiny uses GPU better than Darknet-19 in YOLOv2-tiny, making it more efficient to evaluate and faster [14]. Meanwhile, YOLOv3-tiny training time is seven times faster than YOLOv3. That's because the YOLOv3-tiny network structure simplifies the network structure of YOLOv3.

The ground truth image and the comparison of the detection results of YOLOv3-tiny, YOLOv3, and YOLOv2-tiny can be seen in Fig. 4. The image in row A is a ground truth image that marks the alveolar bone as a yellow rectangle with four green vertices. Meanwhile, the images in rows B, C, and D are the results of detection of alveolar bone using the YOLOv3-tiny, YOLOv3, and YOLOv2-tiny methods which are illustrated in the form of a yellow bounding box with labels and detection confidence scores. In column I and column II, all methods were able to detect alveolar bone correctly. Column I is an example of the detection of one alveolar bone and column II is the detection of two clearly visible alveolar bones. In both detection examples, YOLOv3 had the largest confidence score of all methods, while YOLOv2-tiny had the smallest confidence score. In column I, the detection

confidence score using YOLOv3 tiny was 99% and YOLOv3 was 100%, while using YOLOv2-tiny was 92%. In column II, for the detection of two visible alveolar bones, the confidence scores using YOLOv3-tiny and YOLOv3 were very good above 90%, whereas using YOLOv2-tiny was less than 90%. In column III, the left alveolar bone is only visible in a small part of the image. YOLOv3-tiny and YOLOv3 can still detect it even with a small confidence score. Meanwhile, YOLOv2-tiny couldn't detect it. This is an example of false negative in alveolar bone detection and related to the smallest YOLOv2-tiny recall value of the two comparison methods.

The precision and recall values of YOLOv3 are the best compared to other methods, as shown in Table I. This is because the YOLOv3 architecture uses three scales, so that YOLOv3 is the best at detecting small to large alveolar bones. However, in the dataset used, most of the alveolar bone objects in the dataset are medium or large in size, so that YOLOv3-tiny which uses a two-scale architecture can detect alveolar bone with a mAP value as good as YOLOv3. While the detection results using YOLOv2-tiny are the lowest because the network architecture only uses one scale, so that feature extraction is not as good as the other two methods.
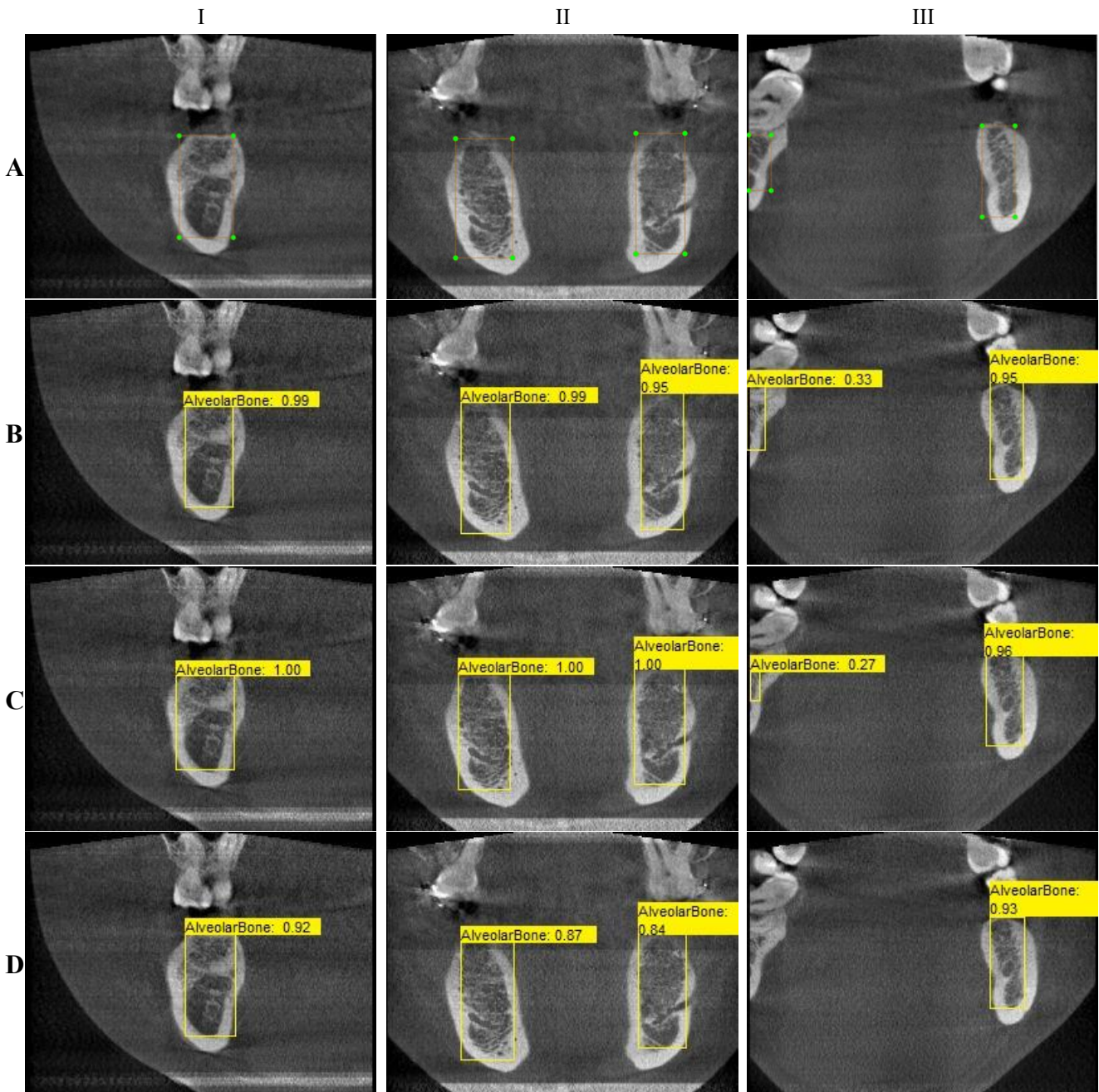
Fig. 4. Alveolar bone detection result. From top to down, test input image (A) ground truth image; (B) YOLOv3-tiny detection result; (C) YOLOv3 detection result; (D) YOLOv2-tiny detection result.

From the results of the experiment, it can be seen that automatic detection of alveolar bone using YOLOv3-tiny for CBCT images can detect alveolar bone with excellent accuracy almost as good as YOLOv3, and outperform YOLOv2-tiny. It is also most efficient in training a detection model for detecting alveolar bone. Automatic alveolar bone detection with excellent results and a fast training process is required to assist dental implant planning.

IV. CONCLUSION

In this study, automatic alveolar bone detection using YOLOv3-tiny for CBCT images resulted in the same mean average precision (mAP) as YOLOv3 of 98.6% and outperformed YOLOv2-tiny with mAP of 96.73%. YOLOv3-tiny training time is faster than YOLOv2-tiny and seven times faster than YOLOv3. The automatic alveolar bone detection system using YOLOv3-tiny based on CBCT images can help the dental implant planning process run more efficiently with accurate results.

When planning dental implants in the posterior mandible, it is necessary to consider the proper distance to the mandibular canal to prevent the risk of alveolar nerve injury. Therefore, for further research, it is necessary to carry out the detection process of the mandibular canal to prevent serious complications. Furthermore, the location of the mandibular canal is required to measure the implant height. Dental implant height is measured based on the distance from the

crest of the alveolar bone and the mandibular canal. Therefore, the location of the mandibular canal and alveolar bone must be detected simultaneously to obtain an accurate implant height.

REFERENCES

[1] G. Juodzbalys and M. Kubilius, "Clinical and radiological classification of the jawbone anatomy in endosseous dental implant treatment," Journal of oral & maxillofacial research, vol. 2, no 2, July 2013.

[2] Y. Mittal, G. Jindal, and S. Garg, "Bone manipulation procedures in dental implants," Indian J Dent, vol. 7, pp. 86-94, 2016.

[3] M. A. F. Banu, and S. Dharman, "Alveolar ridge dimension and morphology measurement in anterior maxilla for immediate implant treatment planning: a Cone Beam Computed Tomography study," Indian Journal of Public Health Research & Development, vol. 10, no. 12, 2019.

[4] R. Indraswari, A. Z. Arifin, N. Suciati, E. R. Astuti, and T. Kurita, "Automatic segmentation of mandibular cortical bone on Cone-Beam CT Images based on histogram thresholding and polynomial fitting," International Journal of Intelligent Engineering & Systems, vol. 12, no.4, Mar 2019.

[5] K. C. T. Nguyen, D. Q. Duong, F. T. Almeida, P.W. Major, N. R. Kaipatur, T. T. Pham, E. H. M. Lou, M. Noga, K. Punithakumar, and L. H. Le, "Alveolar bone segmentation in intraoral ultrasonographs with machine learning," J Dent Res, vol. 99, no. 9, pp. 1054-1061, August 2020.

[6] D. Q. Duong et al., "Fully automated segmentation of alveolar bone using deep convolutional neural networks from intraoral ultrasound images," 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, pp. 6632-6635, 2019.

[7] S. Lee, S. Woo, J. Yu, J. Seo, J. Lee and C. Lee, "Automated CNN-based tooth segmentation in Cone-Beam CT for dental implant planning," in IEEE Access, vol. 8, pp. 50507-50518, 2020.

[8] H. J. Chang, et al. "Deep learning hybrid method to automatically diagnose periodontal bone loss and stage periodontitis." Scientific reports vol. 10 (1): 7531, 5 May 2020.

[9] J. Redmon, S. Diwala, R. Girshick, and A. Farhadi, "You Only Look Once: unified, real-time object detection," ArXiv:1506.02640, 2016.

[10] G. H. Aly, M. Marey, S. A. El-Sayed, and M. F. Tolba, "YOLO based breast masses detection and classification in full-field digital mammograms," Computer Methods and Programs in Biomedicine, vol. 200:105823, Mar 2021.

[11] R. Indraswari, T. Kurita, A. Z. Arifin, N. Suciati, E. R. Astuti, and, A. D. Navastara, "3D region merging for segmentation of teeth on Cone-Beam Computed Tomography images," 2018 Joint 10th International Conference on Soft Computing and Intelligent Systems (SCIS) and 19th International Symposium on Advanced Intelligent Systems (ISIS), pp. 341-345, 2018.

[12] R. H. Putra, N. Yoda, M. Iikubo, Y. Kataoka, K. Yamauchi, S. Koyama, U. Cooray, E. R. Astuti, T. Takahashi, and K. Sasaki, "Influence of bone condition on implant placement accuracy with computer-guided surgery," Int J Implant Dent, 6(1):62, Sep 2020.

[13] A. Bochovskiy, "YOLO v4, v3, and v2 for Windows and Linux," Available online: https://github.com/AlexeyAB/darknet (accessed on 2 December 2020).

[14] W. Fang, L. Wang, and P. Ren, "Tinier-YOLO: a real-time object detection method for constrained environments," in IEEE Access, vol. 8, pp. 1935-1944, 2020.

[15] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," ArXiv:1804.02767, 2018.

[16] Y. Tian, G. Yang, Z. Wang, H. Wang, E. Li, and Z. Liang, "Apple detection during different growth stages in orchards using the improved YOLO-V3 model," Computers and Electronics in Agriculture, vol. 157, pp. 417-426, 2019.