MEDINFO 2021: One World, One Health – Global Partnership for Digital Innovation
P. Otero et al. (Eds.)

doi:10.3233/SHTI220165

Correlation Between Quality Evaluation Metrics and Teeth Detection Results in Panoramic X-Rays Using Deep Learning

Claudia L. Giardina[a], Horacio Legal[a], José Luis Vázquez Noguera[a], Luis Salgueiro[b], Vicente R. Fretes[c], Diego Defazio[d]

[a] Facultad Politécnica, Universidad Nacional de Asunción, Central, Paraguay
[b] Universitat Politècnica de Catalunya, Barcelona, España
[c] Faculdade de Odontologia de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brasil
[c] Facultad de Odontología, Universidad Nacional de Asunción, Asunción, Paraguay
[d] Centro de Diagnóstico y Tratamiento Periodontal, Asunción, Paraguay

Abstract

Panoramic images are one of the most requested exams by dentists for allowing the visualization of the entire mouth. Interpreting X-ray images is a time-consuming task in which misdiagnoses can occur due to the inexperience or fatigue of professionals. In this work, we applied different image enhancement techniques as a pre-processing step to determine which image features correlate with improvements in teeth detection in panoramic images using deep learning architectures. We contrasted the performance of five object-detection architectures using 300 panoramic images of a public dataset. We evaluated the enhancement in the pre-processing step and the detection performance. Quality and detection metrics were considered, and the cross-correlation between them was computed for every object-detection method contemplated. We observe the dependence of the detection performance with some image enhancement techniques, especially those that introduce less noise and preserve the global contrast of the image.

Keywords:

Panoramic images, Deep Learning, Teeth detection

Introduction

Panoramic radiography is an extra-oral type of dental X-ray image that has become commonly used in dental practice and can be a valuable diagnostic tool in the dentist's armamentarium. It allows the visualization of the maxillary and mandibular teeth, surrounding bones, chin, and part of the cervical spine [1], complementing the clinical examination while exposing the patient to a small dose of radiation [1,2]. The interpretation of X-ray images is a time-consuming task mainly performed by dentists where misdiagnosis can occur due to the inexperience, fatigue, or bias of the professional [3]. Hence, a computer-assisted application is a helpful tool for assisting diagnosis and reducing the workload of the professionals.

Teeth detection can be done by Image-Processing and Deep Learning (DL) algorithms. In the field of dentistry, neural networks are being mainly used for the detection and segmentation of teeth, with a few works focusing on the classification of dental pathologies as caries and periodontal bone loss [2]. Recent tooth detection and classification works achieve results comparable to professionals. Tuzoff et al. [4] achieved a sensitivity of 99.41% and a precision of 99.45% using Faster-RCNN [5] architecture for teeth detection in panoramic radiographs. Chen et al. [2] developed three post-processing techniques to improve the Faster-RCNN detections, achieving a precision and sensitivity of 98.8% and 98.5%, respectively. In [6] a fully convolutional neural network based on U-Net [7] is used to perform semantic segmentation on panoramic images, achieving a final accuracy of 95.06%. The success of a DL method not only depends on the configuration of the network (the architecture), but also, is related to other aspects like pre-processing, data augmentation, and hyper-parameter optimization. In this work, we explored some pre-processing, which has proven to be a differentiating factor between DL algorithms with the same architecture [8].

The rest of the article is organized as follows: in Methods we present the description of the dataset, the Image Enhancement Techniques (IET), the object detection architectures, and the experimental setup. In Results, we show the object detection performance of each architecture, as well as the correlation between the detection results and the Image Quality Metrics (IQM). Finally, we present the discussion and conclusion remarks.

Methods

Dataset

We use a public dataset of 1500 panoramic images generated by Gil Silva et al. [9] originally conceived for semantic segmentation. From the 1500 images, 300 images were selected and divided into 192 for training, 48 for validation, and 60 for testing. Bounding boxes were drawn for each tooth, implant, prosthesis, or dental root fragments as the ground-truth labels. These annotations were reviewed, edited, and validated by two experienced dentists.

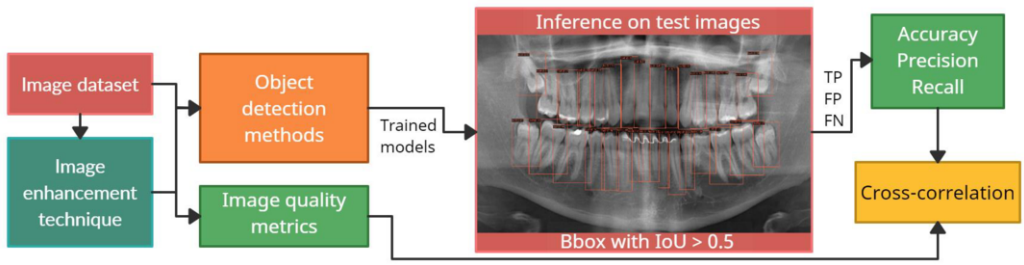

Figure 1– Diagram of steps followed in this work.

Table 1– Hyperparameters values.

Architecture	batch size	learning rate	weight decay	warmup iterations	steps
Faster R-CNN	2	0.01	0.0001	48	6
RetinaNet	2	0.008	0.0001	4	4
Cascade R-CNN	2	0.009	0.001	24	3, 5
FCOS	2	0.02	-	48	6, 13
$YOLOv_3$	8	0.001	0.001	-	48, 96

Image enhancement techniques (IET)

The image enhancement techniques considered were the smoothing-edge preservation filters Anisotropic Diffusion (AD) [10] and Bilateral Filtering (BF) [11], in addition to the contrast enhancement algorithms BBHE [12], CLAHE [13], GRMMCE [14], MTHT [15], and QHELC [16]. Before applying the mentioned techniques, the images were first trimmed to exclude extra toothless regions by having as references the hard palate and lower mandibular contour. For evaluating the image enhancement, we considered quality metrics like Absolute Mean Brightness Error (AMBE), Peak Signal to Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), standard deviation or global contrast (GC), Contrast Improvement Ratio (CIR) [17], Entropy, and Edge Preservation Index (EPI) [18].

AMBE measures if a method preserves the original brightness of an image. The better the brightness preservation, the lower AMBE would be. PSNR measures the distorting noise introduced in an image, where a higher image quality yields a higher PSNR. SSIM measures the structural similarity between the original image and the output image with a better value closer to 1. CIR measures the enhancement in the local contrast of the images. A higher value of CIR indicates a higher difference between the local contrast of the original image and the output image [17]. Entropy is used to quantify the information content of an image; a more detailed image has the higher entropy. EPI index quantifies the edge preservation after applying the enhancement to the image. EPI equals 1 indicates a similar value between the input and enhanced images [18].

Object detection architectures

We selected Faster R-CNN [5], Cascade R-CNN [19], RetinaNet [20], Fully Convolutional One-Stage Object Detection (FCOS) [21], and You Only Look Once (YOLOv3) [22] for being among the most widely used architectures. Pre-trained models were chosen from the Model Zoo of MMDetection [23], an open-source object detection toolbox based on Pytorch. Backbones used were ResNext [24] for Faster R-CNN, RetinaNet, and Cascade R-CNN, ResNet 101 for FCOS, and DarkNet-53 for YOLOv3. All pre-trained models were trained on the COCO dataset.

The best weight of each method, trained with the panoramic images, was chosen considering the highest mean Average Precision (mAP) with $IoU \geq 0.5$, The IoU is defined as: $IoU = \frac{Area_{DB} \cup Area_{GTB}}{Area_{DB} \cap Area_{GTB}}$, where $Area_{DB}$ is the area of detected boxes and $Area_{GTB}$ is the area of ground truth boxes.

With the models trained and the weights selected, the performance of the models on the test set was evaluated considering as true detection a prediction with $IoU \geq 0.5$. We then measured the accuracy (ACC), precision (PRE), and recall, metrics that are calculated using: $Accuracy = \frac{TP}{TP+FP+FN}$, $Precision = \frac{TP}{TP+FP}$, and $Recall = \frac{TP}{TP+FN}$, where TP or True Positive is the number of teeth correctly detected, FP or False Positive is the number of teeth incorrectly detected, and FN or False Negative is the number of teeth that are not detected or are detected with and $IoU < 0.5$. Figure 1 shows a diagram of the steps followed in this work.

Experimental setup

Table 1 shows the parameters used for training the models, these values were set after hyperparameter tuning. The baseline model was trained using the original dataset, and seven more models were trained with the dataset obtained after the application of different IET. We chose the best weight for each architecture, after training for at most 100 epochs.

Results

Image quality metrics (IQM)

Table 2 shows the IQM for the pre-processed images as well as the Entropy and GC measured on the original dataset. We noticed that the IET that introduces less noise to the images was the bilateral filtering, achieving the highest PSNR among the techniques. While the change in the contrast made by BBHE,

introduces the most distortion and noise to the images, having low values of SSIM and PSNR. CLAHE also introduces noise to the images, but unlike BBHE, CLAHE preserves the edges of the objects, having the closest value of EPI to one, while also raising the entropy of the images. Applying QHELC produces an image with an EPI close to one, which also has the highest SSIM value, indicating a low distortion introduced to the images.

Teeth detection

Table 3 shows the detection performance of the methods. Faster R-CNN results show that five models have a better accuracy than the baseline model, being the MTHT model the one with the best results, and the BF model the second best. On the contrary, QHELC model is the one with the lowest detection performance, having the worst accuracy and precision among the models. CLAHE is the model with the highest recall across all models.

In RetinaNet results, we noticed that GRMMCE, bilateral filtering, and QHELC models have better accuracy than the baseline. Being GRMMCE the one with the best accuracy, whereas BBHE has the lowest accuracy and precision among the models.

According to Cascade R-CNN results, QHELC model is the only model that has a slightly better performance than the baseline model, while the model with the worse performance is again the BBHE model, having a precision 5.37% less than the precision of the QHELC model.

In FCOS results, the accuracy and recall have the same value since the precision is 100%. For this architecture, QHELC model achieves the higher accuracy, and it is followed by the baseline model, while the BBHE model presents the lowest accuracy among the models.

Finally, in YOLOv3 detection results, we noticed that there were no improvements introduced by any IET. Conversely, the BBHE model presents the lowest accuracy among the models.

Cross-correlation

We computed the correlation heat maps presented in Figure 2 using the IQM in Table 2 and the detection results of Table 3.

Figure 2-A shows the correlation heat map of Faster R-CNN. The relationships between the detection metrics and the IMQ are mostly neutral in this heat map. The relationship that slightly stands out is the negative relation of the precision with AMBE and GC. RetinaNet models' correlation heat map is shown in Figure 2-B. The map shows that the precision has a negative relationship of -0.27 with GC, and -0.22 with CIR. In Figure 2-C, the heat map of Cascade R-CNN models shows that the accuracy and the precision have a negative relation with AMBE, the GC, and CIR. While they have a positive relationship with both SSIM and PSNR. The heat map of FCOS in Figure 2-D does not have the precision since the value of this metric is 100% for every model. According to the heat maps in Figure 2-D and 2-E, the teeth detection results of FCOS and YOLOv3 models have a neutral relationship with the IQM considered.

Table 2 – Average of IQM measured on the testing group.

	AMBE	Entropy	SSIM	PSNR	EPI	Contrast	CIR
Original dataset		6.998				33.120	
AD	9.4E-07	6.982	0.963	41.003	0.423	32.715	0.742
BF	2.2E-06	6.987	0.970	42.996	0.448	32.863	0.680
BBHE	5.7E-02	6.829	0.710	14.810	0.423	74.154	39.204
CLAHE	5.9E-02	7.647	0.866	19.038	0.951	50.484	1.699
GRMMCE	5.2E-04	7.058	0.959	35.632	0.672	34.578	4.356
MTHT	6.0E-04	7.156	0.767	27.793	0.309	37.119	35.199
QHELC	1.1E-03	6.982	0.993	36.906	0.913	36.334	0.188

Table 3– Teeth detection performance of the methods.

	Faster R-CNN			RetinaNet			Cascade R-CNN			FCOS		YOLOv3		
	ACC	PRE	Recall	ACC	PRE	Recall	ACC	PRE	Recall	ACC	PRE	ACC	PRE	Recall
Baseline	98.92	99.49	99.43	98.81	99.43	99.37	99.43	99.94	99.49	97.83	100	96.29	99.94	96.34
AD	99.09	99.66	99.43	98.69	99.37	99.31	99.37	99.89	99.49	97.43	100	94.18	99.82	94.34
BF	99.20	99.71	99.49	98.92	99.43	99.49	99.37	99.94	99.43	97.54	100	95.03	99.88	95.14
BBHE	98.81	99.32	99.49	97.37	97.92	99.43	95.35	95.76	99.54	97.37	100	95.84	99.76	96.05
CLAHE	99.15	99.49	99.66	98.70	99.09	99.60	97.75	98.25	99.49	97.71	100	95.66	99.82	95.83
GRMMCE	99.03	99.54	99.49	98.98	99.43	99.54	99.37	99.83	99.54	97.54	100	95.95	99.82	96.11
MTHT	99.26	99.71	99.54	98.41	99.03	99.37	99.31	99.94	99.37	97.71	100	95.04	99.64	95.37
QHELC	98.64	99.31	99.31	98.92	99.37	99.54	99.49	99.94	99.54	98.11	100	95.83	99.94	95.88

Discussion

The detection performance of the models, shown in Table 3, indicates that applying an enhancement to the dataset does not always result in an improvement of the detection. This behavior is found in Cascade R-CNN, FCOS, and YOLOv3, where the baseline model has the second best or the best result. We can also observe that the QHELC model achieves the best results in Cascade R-CNN and FCOS, and the second-best result in RetinaNet. The BBHE model shows the weakest or one of the weakest performances in the five architectures. The heat map with the strongest relationships belongs to Cascade R-CNN in Figure 2-C. The negative relationship of the precision with GC and CIR, also presented in the RetinaNet heat map, means that the precision will be negatively affected by an increment of GC and local contrast. According to Table 2, BBHE is the enhancement technique that increases the most GC and the local contrast of the images, which results in the BBHE model having the lowest precision among the models of RetinaNet and Cascade R-CNN. We can see this in Figure 3, where in A we have the inference of the original model of RetinaNet, and in B we see the BBHE model. In Figure 3-A, a fragment of a tooth is missed, while in Figure 3-B the fragment is detected but there are two extra boxes, these boxes affect the precision of the model. The heat map of Figure 2-C also shows that the accuracy and the precision have a positive relationship with SSIM and PSNR. This means that techniques that less noise and distortion introduced to the images, like QHELC, have a greater chance to have higher accuracy and precision.

The correlations found in this work might vary if the dataset has a higher number of images, or with greater diversity in the features of the images. For that reason, increasing the number of images in the dataset would elevate the importance of the results.

Conclusions

We evaluated the detection performance of several deep learning object-detection architectures with images modified by seven IET as a pre-processing step in the context of teeth detection. Some models show dependency on the modification of the image while others do not. We also correlated the features obtained by the IET with improvements of the detection performance showing that images with the highest values of SSIM and PSNR, i.e., those IET which preserves better the structure and introduce less noise to the images reach better performance and are suitable to be used for data augmentation, while incrementing CIR and GC will be in detrimental of the detection performance.

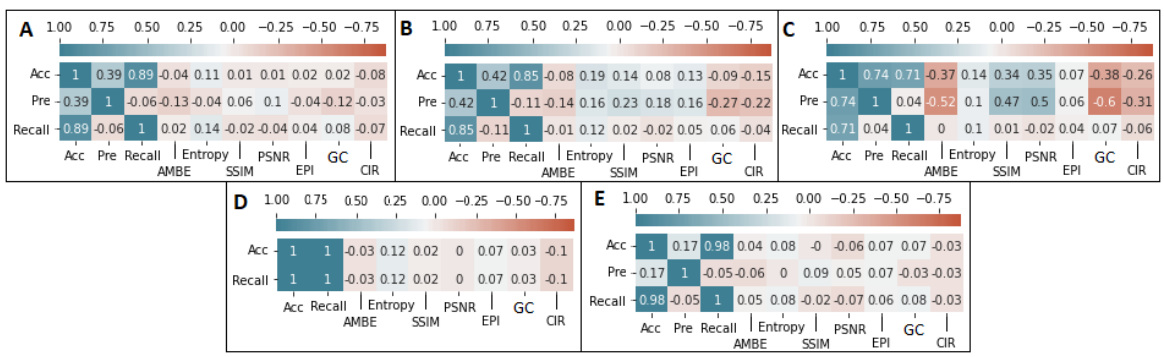

Figure 2– Correlation heat maps of A: Faster R-CNN, B: RetinaNet, C: Cascade R-CNN, D: FCOS, and E: YOLOv3.

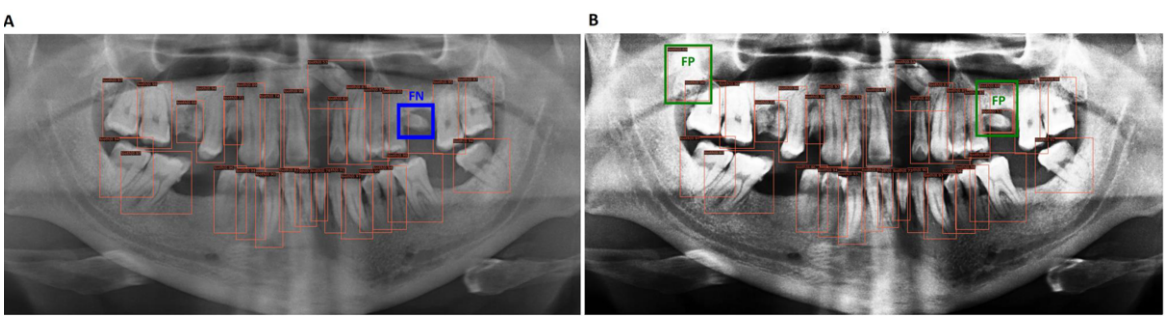

Figure 3– A: Baseline model inference missed a tooth marked with a blue box. B: BBHE model inference has two extra boxes marked with the green boxes.

References

[1] G. Jader, J. Fontineli, M. Ruiz, K. Abdalla, M. Pithon, and L. Oliveira, Deep instance segmentation of teeth in panoramic x-ray images, In Conference on Graphics, Patterns and Images (SIBGRAPI) (2018), 400-407.

[2] F. Schwendicke, T. Golla, M. Dreher, and J. Krois, Convolutional neural networks for dental image diagnostics: A scoping review, Journal of Dentistry 91 (2019), 103226.

[3] H. Chen, K. Zhang, P. Lyu, H. Li, L. Zhang, J. Wu, and C.-H. Lee, A deep learning approach to automatic teeth detection and numbering based on object detection in dental periapical films, Scientific Reports 9 (1) (2019).

[4] D. Tuzoff, L. Tuzova, M. Bornstein, A. Krasnov, M. Kharchenko, S, Nikolenko, M. Sveshnikov, and G. Bednenko, Tooth detection and numbering in panoramic radiographs using convolutional neural networks, Dentomaxillofacial Radiology 48 (4) (2019), 20180051.

[5] S. Ren, K. He, R. Girshick, and J. Sun, Faster r-cnn: Towards real-time object detection with region proposal Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6) (2015).

[6] T. Koch, M. Perslev, C. Igel, and S. Brandt, Accurate segmentation of dental panoramic radiographs with u-nets, In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019) (2019), 15-19.

[7] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, LNCS 9351 (2015), 234-241.

[8] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, A survey on deep learning in medical image analysis, Medical Image Analysis 42 (2017), 60-88.

[9] G. Silva, L. Oliveira, and M. Pithon, Automatic segmenting teeth in x-ray images: Trends, a novel data set, benchmarking and future perspectives, Expert Systems with Applications 107 (2018), 15-31.

[10] G. Sapiro and D. Ringach, Anisotropic diffusion of multivalued images with applications to color filtering., IEEE Transactions on Image Processing 5 (1996), 1582-1586.

[11] C. Tomasi and R. Manduchi, Bilateral filtering for gray and color images, In Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271) (1998), 839-846.

[12] Y.-T. Kim, Contrast enhancement using brightness preserving bi-histogram equalization, IEEE Transactions on Consumer Electronics 43 (1997), 1-8.

[13] K. Zuiderveld, Contrast Limited Adaptive Histogram Equalization, Graphics Gems, 4th Edition, Elsevier, 1994, 474-485.

[14] J. Mello, R. Escobar, F. Martínez, J. Vázquez Noguera, H. Legal-Ayala, and D. Pinto-Roa, Medical image enhancement with brightness and detail preserving using multiscale top-hat transform by reconstruction, Electronic Notes in Theoretical Computer Science 349 (2020), 69-80.

[15] J. Mello, H. Legal-Ayala, and J. Vázquez Noguera, Image color contrast enhancement using multiscale morphology, Computational Interdisciplinary Sciences 8 (3) (2017).

[16] I. Brizuela, R. D. Medina Caballero, J. Silva, J. Roman, and J. Vazquez Noguera, Quadri-histogram equalization using cutoff limits based on the size of each histogram with preservation of average brightness, Signal, Image and Video Processing 13 (2) (2019).

[17] Y.-P. Wang, Q. Wu, K. Castleman, and Z. Xiong, Chromosome image enhancement using multiscale differential operators, IEEE Transactions on Medical Imaging 22 (5) (2003), 685-693.

[18] F. Sattar, L. Floreby, G. Salomonsson, and B. Lovstrom, Image enhancement based on a nonlinear multiscale method, IEEE Transactions on Image Processing 6 (6) (1997), 888-895.

[19] Z. Cai and N. Vasconcelos, Cascade R-CNN: high quality object detection and instance segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (2019), 1483-1498.

[20] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, Focal loss for dense object detection, IEEE Transactions on Pattern Analysis and Machine Intelligence (2017), 1-1.

[21] Z. Tian, C. Shen, H. Chen, and H. Tong, Fcos: Fully convolutional one-stage object detection, In 2019 IEEE/CVF International Conference on Computer Vision (ICCV) (2019), 9626-9635.

[22] J. Redmon and A. Farhadi, Yolov3: An incremental improvement, (2018).

[23] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, J. Shi, W. Ouyang, C. C. Loy, and D. Lin, MMDetection: Open mmlab detection toolbox and benchmark, (2019).

[24] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, Aggregated Residual Transformations for Deep Neural Networks, In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017), 5987-5995.

Address for correspondence

Claudia L. Giardina, email address cgiardina@fpuna.edu.py.