BC3409: AI IN ACCOUNTING & FINANCE

Group Project Report

Segmentation of Teeth to Facilitate Dental Diagnosis

SEMINAR GROUP 1 TEAM:

Group Member	Matriculation Number
Aileen Laksmono Lie	U1920118E
Chua Cheng Hong	U1721774L
Dimas Valls Quiros	N2202617J
Maelys Jordane Boudier	N2202193F
Rajkumar Snehaa	U1921000H
Timmothy Yonathan	U2040650E

Table of Content

Table of Content 2

Executive Summary 4

1. Problem Statement 5

- 1.1 Introduction 5
- 1.2 Objective 5

2. Analytical Approach 6

- 2.1 Instance Segmentation Problem 6
- 2.2 Model Training and Evaluation 7

3. Data Set 7

- 3.1 Dataset 7
- 3.2 Data Annotation 7
- 3.3 Data augmentation 8

4. Technology Used 9

- 4.1 MMDetection Model Structure 9
- 4.2 Mask R-CNN 10
- 4.3 SOLOv2 11

5. Methodology 13

- 5.1 Model Parameters 13
  - 5.1.1 Mask-R-CNN 13
  - 5.1.2 SOLOv2 13
- 5.2 Evaluation Set-Up 14
  - 5.2.1 Training 14
  - 5.2.2 Hyperparameter Tuning 14
  - 5.2.3 Evaluation 15

6. Results 15

7. Business Implications 16

- 7.1 Direct Benefits of AI models 16
- 7.2 Benefits of Instance Segmentation vs Semantic Segmentation 17
- 7.3 Benefits of Segmenting Teeth vs Direct Diagnosis 17

8. Limitations 17

- 8.1 Dataset Characteristics 17

9. Future Work 18

- 9.1 Improving the model 18

9.2 Business Opportunity: Telemedicine 18
10. Conclusion **19**
References **20**

Executive Summary

This study focuses on creating a model capable of segmenting teeth to facilitate dental diagnosis. With proper teeth segmentations, future models will be able to detect anomalies and combine the segmentation with various use cases such as telemedicine and treatment recommendations. This report begins with the problem statement and objectives, followed by our analytical approach, the dataset and technology used, our methodology, and finally our results and insights.

Breakthroughs in Deep Learning and AI have greatly impacted healthcare. The sensitivity of data can be somewhat of a pushback as datasets are hard to come by and must be treated with care. However, investigating how technology can be used to facilitate the diagnosis of medical anomalies can significantly advance the diagnostics landscape, which can prove to be very impactful in underdeveloped areas where medical access is limited.

To conduct our research, we worked in partnership with TrueVA which aims to use computer vision in detecting and labeling teeth from intraoral photographs. They provided a dataset of 500 images which we labeled by drawing polygons around the teeth. We repeated this process on a smaller public dataset to run our model and provide some key results in this report.

While our team managed to label the 500 proprietary images owned by TrueVA on their devices, there was a major setback in developing the models on their computers given the lack of computational resources (and we could not use Google Colab to ensure data privacy). However, we found a much smaller public dataset with similar photos as those provided by TrueVA upon which we were able to develop the models.

This study focuses on the use of instance segmentation techniques; more specifically, we used Mask-R-CNN and SOLOv2. We cover the application, strength and use cases of both convolutional neural networks to demonstrate their purpose. Finally, we discuss business implications and limitations from a more global perspective.

Deep learning models serve a greater purpose than simple teeth segmentation as they will revolutionize healthcare. Machines lack the ‘human bias’ that doctors may have and can’t get tired after a long day’s work. As such, they can greatly help in improving medical diagnosis. We hope that the rollout of our model will prove beneficial to our partnering dentist’s office and that it can be expanded to more medical cabinets across Singapore.

1. Problem Statement

1.1 Introduction

The improvement in quality of life over the years has brought about more attention to dental health, as it is one of the integral parts of the daily functioning of the human body. The structure of an individual's teeth affects the process of digestion, while their appearance influences other people's perceptions and contributes to one's social status in society. This has led to a rise in the popularity of orthodontic treatment, where individuals born with suboptimal dental features are able to ameliorate their teeth's appearance.

Successful orthodontic treatment requires an accurate diagnosis that leads to an optimized treatment plan as a first step. Oftentimes, an orthodontist collects various types of patient data before making a sound judgment. These data can be in the form of X-ray radiographs, extra- and intra-oral photographs, dental models, etc. (Im et al., 2022).

One of the first things a dentist does when screening for a patient's problems on these data is to analyze teeth shape, number, and position. Traditionally, this process was conducted solely with the trained eyes of dentists. This can be time-consuming and error-prone as it is manual and requires a high caliber of professional dentistry aptitude.

1.2 Objective

In recent years, the advancement of AI has made computer-aided diagnosis (CAD) become an important tool for automation in dentistry. Through its implementation and due to its credibility as a robust tool trained on thousands to millions of data points, CAD provides dentists with a trustworthy initial diagnosis, consequently saving time, reducing the impact of fatigue from daily practice in clinical decision-making and improving accuracy of diagnosis (Tuzoff et al., 2019).

In this project, we have collaborated with TrueVA, whose ultimate goal is to derive a CAD tool that will provide an initial diagnosis of the severity of the patient's problems. The objective of our collaboration is to develop a robust AI model that will automate teeth numbering according to the Federation Dentaire Internationale(FDI) Tooth Numbering System shown in Fig. 1 below. The developed model will then be used as a crucial stepping stone to deriving the severity of dental issues faced by the patient using subsequent mathematical pre-processing by TrueVA.

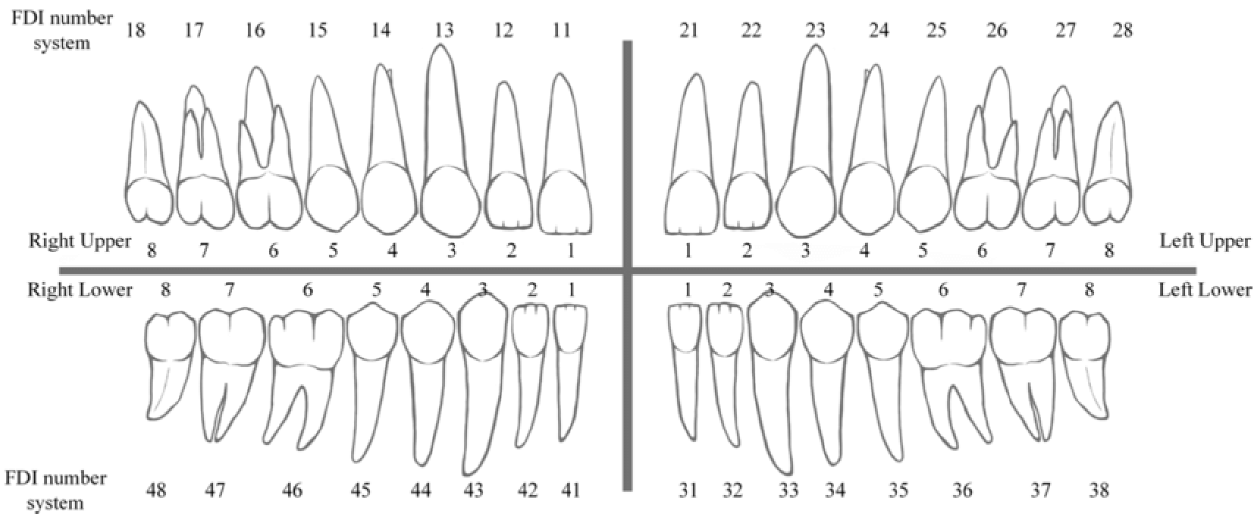


Fig. 1: FDI Tooth Numbering System

2. Analytical Approach

Fig. 2 below highlights the overarching analytical steps taken to achieve our objective.

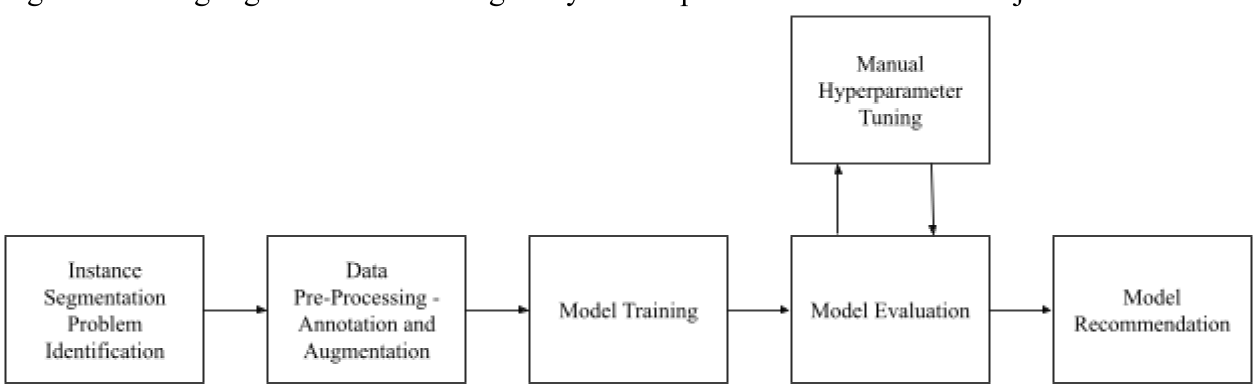


Fig. 2: Overview of Analytical Approach

2.1 Instance Segmentation Problem

As identifying just the bounding boxes of each tooth will limit the amount of dental problems that can be identified down the line, there is a necessity to identify the boundaries of the individual teeth at the pixel level. Therefore, through analyzing our objective, we have come to the conclusion that the problem we will be addressing with our recommendations is an instance segmentation problem.

Instance segmentation is a special form of image segmentation that focuses on detecting instances of objects and demarcating their specific boundaries beyond simple bounding boxes. It is a combination of object detection, which allows detection of many objects from different classes in each image, and image segmentation, that makes a pixel-level prediction where each pixel in an image is classified according to a category.

Through identifying our problem statement to be an instance segmentation problem, we will be able to pick suitable models and appropriately evaluate them with regards to the demands of the problem.

2.2 Model Training and Evaluation

We have decided to implement and evaluate the performance of two state-of-the-art models, namely Mask R-CNN and SOLOv2, that are used for image segmentation. We have decided to use MMdetection, an object detection toolbox for training and evaluation of our chosen instance segmentation problems. The mechanisms of the models chosen will be explained in greater detail in Section 4.

3. Data Set

3.1 Dataset

TrueVA provided a dataset of 500 intraoral dental images, consisting of 300 teeth-together and 200 teeth-apart photographs. Additionally, we sourced another public dataset sourced from *Oral and Dental Spectral Image Database* (ODSI-DB). It contains 52 intraoral dental photographs which were used to generate the teeth segmentation models.

3.2 Data Annotation

To ensure that the models learn to identify the teeth shape, number, and position correctly, the dataset was annotated beforehand according to the FDI Numbering System. Bounding boxes were drawn over 16 front teeth in each of the images: 8 on the upper jaw and 8 on the lower jaw. To do this, the team used Makesense.ai, which is a free online tool for labeling photos in computer vision projects. An illustration is provided below in Fig. 3.1.

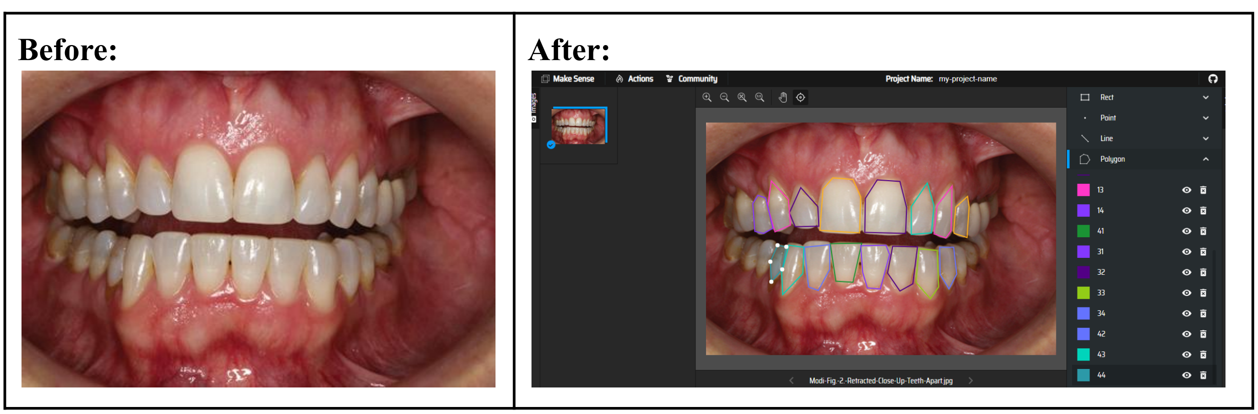


Figure 3.1: Annotation Example

3.3 Data augmentation

Image augmentation can improve the generalizability of the model's prediction by increasing the variety of the image dataset that the model was trained on. This is done through minor random changes applied to the dataset, such as random rotation, random cropping, random noise, etc. that effectively increase the sample size.

While creating two versions of the same photo would not be as good as having 2 separate photos of the object, image augmentation is a powerful technique to generate more data from existing ones. In this project, the team decided to use [Roboflow](), a popular tool that eases computer vision tasks, to augment the dataset.

The augmentation steps were applied with consideration to the realistic probability of such images being submitted during actual implementation, so as to enable our model to perform well despite suboptimal image inputs. The base dataset was augmented in the following ways:

1. Rotating the image between -15º and +15º of the original image as highlighted in Fig. 3.2. Rotation was chosen as an augmentation step as patients are likely to take ill-positioned pictures if they are unsure of how the application operated.

Original Image	**-15º**	**+15º**

Figure 3.2: Example of Data Augmentation via Rotation

2. Changing the exposure of the original image between -25% to + 25% as highlighted in Fig. 3.3. Exposure was chosen as an augmentation step as there is the possibility patients take pictures in bad lighting.

Original Image	**-25%**	**+25%**

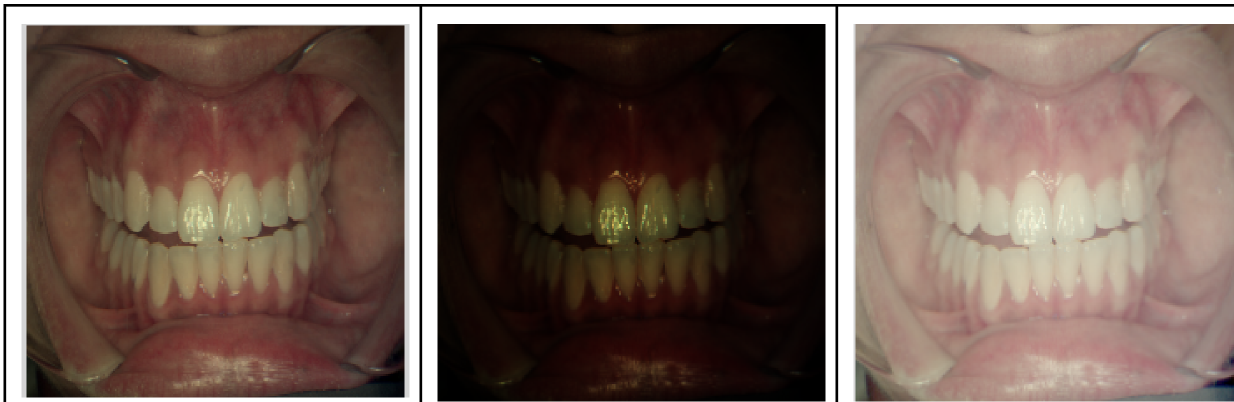


Figure 3.3: Example of Data Augmentation via Rotation

3. Blurring the image up to 2.5px as highlighted in Fig. 3.4. Blurring was chosen as an augmentation step as there is the possibility patients could be taking pictures with less sophisticated cameras.

Original Image	**2.5px**
	

Figure 3.4: Example of Data Augmentation via Rotation

4. Technology Used

4.1 MMDetection Model Structure

The detector models MMdetection toolbox adopts the popular paradigm for deep learning-based object detectors:

- Backbone: The backbone network, usually a Fully Convolutional Network, converts the input image into raw feature maps
- Neck: The neck receives the raw feature maps from the backbone network and enhances the multi-scale features through refinements and reconfigurations.

- Detection Head: After the input image processed by the backbone and neck parts, MMdetection passes the results to detection heads to perform specific tasks such as bounding box prediction and mask prediction.

The modular design of the components in both toolboxes makes it very easy to construct a customized detection framework that best fits our particular use case.

4.2 Mask R-CNN

R-CNN (Region-Based Convolutional Neural Network) is a type of machine learning model that is used for instance segmentation problems. R-CNN utilizes bounding boxes across the object regions, which then evaluates convolutional networks independently on all the Regions of Interest to classify multiple image regions into the proposed class. (Zhang, 2022).

Mask R-CNN is an extension of Faster R-CNN and it has an architecture as shown in Fig 4.1.

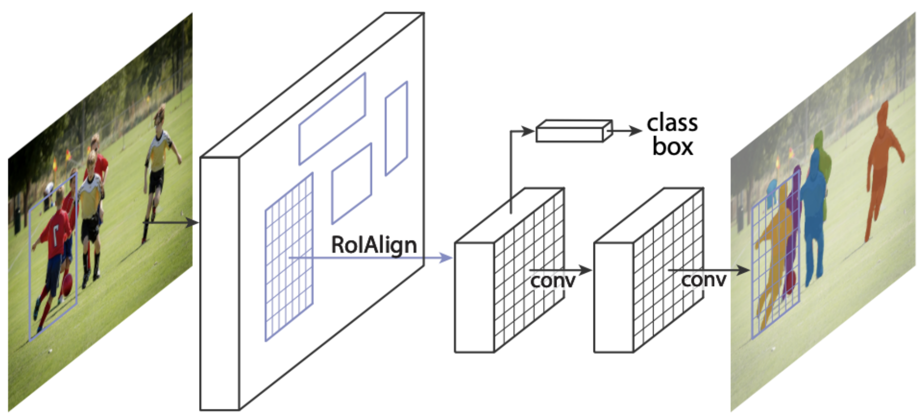


Figure 4.1: Mask R-CNN Architecture (He et al., 2018)

There are two stages to Mask R-CNN,

1. **Stage 1: Region proposals.** The first stage consists of two networks, backbone and region proposal networks. These networks run once per image to give a set of region proposals, which are regions in the feature map that contain the object.
2. **Stage 2: Bounding box recognition and prediction of object class.** The network predicts bounding boxes and object class for each of the proposed regions obtained in stage 1. There is also a **branch for predicting an object mask** in parallel with the existing branch for bounding box recognition. (Khandelwal, 2019).

Model Output
The model produces the output in three components:

- The **bounding boxes** — x1, y1, width, height if using the COCO file format
- The **class** of the bounding box
- The **probability score** for that prediction— how certain the model is that the class is actually the predicted class

Strengths of Mask-R-CNN

- Mask R-CNN is simple to train.
- Mask R-CNN outperforms all existing, single-model entries on every task.
- The method is very efficient and adds only a small overhead to Faster R-CNN.
- Mask R-CNN is easy to generalize to other tasks.

4.3 SOLOv2

SOLO model
SOLO (segment objects by locations) is a simple and flexible framework applied for accomplishing instance segmentation. It assigns each pixel within an instance of an object to a category based on its location and size. *(Wang & Li, 2020).* SOLO formulates the task of instance segmentation as two sub-tasks of pixel-level classification, solvable using standard Fully Connected Networks (FCNs), thus simplifying the formulation of instance segmentation. It takes an image as input, and outputs instance masks and corresponding class probabilities. Its architecture is seen in Figure 4.2.

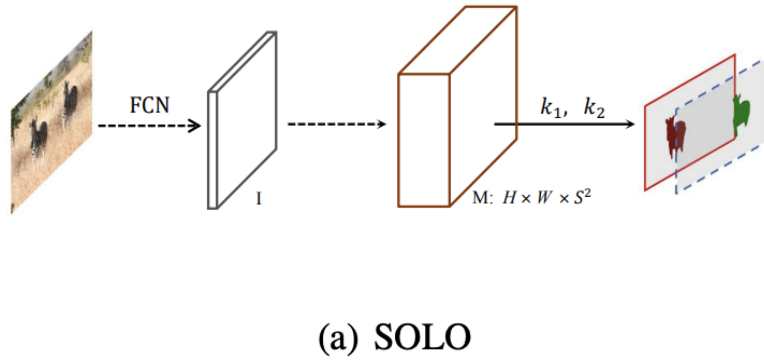


Figure 4.2 : SOLO model Architecture (Wang & Li, 2020)

SOLOv2 model

SOLO is a dynamic scheme for segmenting objects by locations. It divides the mask learning process into two parts, namely convolution kernel learning and feature learning. *(Shiledarbaxi, 2021)*. It assigns appropriate location categories to different pixels while predicts mask kernels dynamically according to the input. It then constructs a unified and high-resolution mask feature representation for instance-aware segmentation. Its architecture can be seen in Figure 4.3.

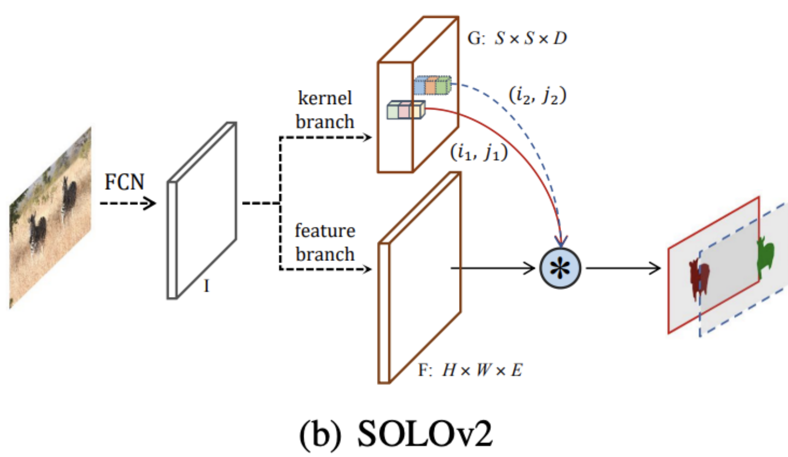


Figure 4.3 : SOLO model Architecture (Wang & Li, 2020)

Model Output

The model produces the output in three components:

- The **bounding boxes** — x1, y1, width, height if using the COCO file format

- The **class** of the bounding box
- The **probability score** for that prediction— how certain the model is that the class is actually the predicted class

Strengths of SOLOv2

- SOLO is not restricted by box locations and scales and hence benefits from the inherent advantages of Fully Connected Networks (FCNs).
- SOLO takes an image as input, directly outputs instance masks and the corresponding semantic class probabilities in a fully convolutional, box-free and grouping-free paradigms

5. Methodology

5.1 Model Parameters

In this section, we provide an overview of the parameters characterizing the models chosen for instance segmentation.

5.1.1 Mask-R-CNN

Parameter	Value
Backbone	Pytorch-implemented Resnet50
Neck	Feature Pyramid Network
Stacked Convolutional Layers	4
Learning Rate	0.001

Table 1: Mask-R-CNN parameters

We chose to make use of the Mask-R-CNN base detector applied with a Feature Pyramid Grid structure and a PyTorch-implemented Resnet50 backbone. Table 1 provides an overview of the important parameters that distinguish the model from its variants.

5.1.2 SOLOv2

Parameter	Value
Backbone	Pytorch-implemented Resnet50
Neck	Feature Pyramid Grid
Stacked Convolutional Layers	4

Parameter	Value
Learning Rate	0.001

Table 2: SOLOv2 parameters

We chose to make use of the SOLOv2 base detector applied with a Feature Pyramid Grid structure and a PyTorch-implemented Resnet50 backbone. Table 2 provides an overview of the important parameters that distinguish the model from its variants.

5.2 Evaluation Set-Up

5.2.1 Training

The distribution of the dataset used during training is shown in Table 3 below.

Category	Number	Percentage
Train	88	70%
Test	38	30%
Total	126	100%

Table 3: Details of Training and Testing Dataset

We split our data into a train and test set with a ratio of 70-30. The training data were used to train the computer vision models, while the test data were used to evaluate the performance of the models on new and unseen data.

5.2.2 Hyperparameter Tuning

We focused on manually tuning 2 parameters, number of epochs and classification loss metrics, to improve the performance of the model. These parameters were chosen due to their important impact on the performance of the model, as explained below:

- **Number of epochs**
    - The number of epochs refers to the number of times the model propagates through the entire dataset during training.
    - Utilizing the right number of epochs during training is important as it minimizes the likelihood of the model overfitting on training data and ultimately failing to generalize well on unseen data.
- **Classification loss function**
    - Classification loss function refers to the loss function that is used to calculate the correctness of the classification of each predicted bounding box during training.

- Utilizing the right classification loss function is important as it is crucial in defining the optimization surface and the decision boundary towards which the algorithm trains to optimize towards.
- We trained the models using both Focal Loss and Cross-Entropy Loss in order to compare performance during evaluation.

5.2.3 Evaluation

We have decided to use average precision and average recall to evaluate the models created.

- **Average Precision** is the weighted average quality of positive predictions by the model
  - Precision is an important metric for our use case as it will impact the dental problem identified down the line and could impact the possibly costly treatments that the patient could be recommended.
- **Average Recall** is the weighted average of the number of correct positive predictions made out of all positive predictions that could have been made by the model.
  - As identification of dental problems usually does not focus only on one tooth but weighted by the combination of different teeth, it is important that the model is able make a large number of correct positive predictions per data record.

6. Results

Average Precision (%)				
	8 Epochs		**16 Epochs**	
	Cross Entropy Loss	*Focus Loss*	*Cross Entropy Loss*	*Focus Loss*
Mask-R-CNN	84.9	85.3	83.9	85.0
SOLOv2	79.6	80.1	85.1	86.8

Table 4: Average Precision of Tested Models

Average Recall (%)				
	8 Epochs		**16 Epochs**	
	Cross Entropy Loss	*Focus Loss*	*Cross Entropy Loss*	*Focus Loss*
Mask-R-CNN	84.9	85.6	84.4	85.1
SOLOv2	80.2	82.0	85.3	86.8

Table 5: Average Recall of Tested Models

In addition to the average precision and recall figures reported in Table 4 and 5, the visualisation of classification loss is also depicted in Table 6 in the Appendix. The best performing model is the Solov2 detection model trained with 16 epochs and a Focus Loss function as it has the highest Average Precision of 86.8% as seen in Table 4 and Average Recall of 86.8% as seen in Table 5. Hence, we will be proceeding with this model, with the specified hyperparameters, for our recommendations to TrueVA.

7. Business Implications

7.1 Direct Benefits of AI models

The medical field generates many images to keep track of the evolution of patients' health and to allow multiple doctors to look at the same pictures. In dentistry, these images - especially once segmented by teeth - can be used in the diagnosis of overbite and underbite, caries detection and other dental anomalies. Deep Learning models have evolved dramatically over the last 30 years and have been applied to the field of medicine. Medical diagnosis is based on the doctor's ability to recognize symptoms and detect anomalies while attributing them to the proper condition (myHSN, 2022). Traditional techniques using computer based tools relied on feature engineering and on the knowledge of field experts, but with machine learning, the model can automatically learn complex features from the original data. Deep learning models now have the ability to parse through dirty data - in our case images - and to create complex structures to help doctors improve diagnosis accuracy. Data scientists can also contribute to the field despite having limited industry knowledge while applying computer vision techniques. As such, we were able to make a detection model to outline the teeth despite not studying dentistry.

The images in our dataset were taken by dentists, but patients could take very similar pictures at home. If the code was integrated into a user-friendly interface, it would allow patients to regularly check up on their teeth at home and determine when a visit to the dentist is necessary for treatment. In countries like the USA where healthcare is very expensive, it could limit trips to the dentist whenever is necessary and allow users to conduct automatic AI-led check-ups online (Peter G. Peterson Foundation, 2022).

Computer vision techniques can also be applied to all other sectors of medicine to aid in diagnosis. For instance, it can be used on x-rays images to detect patterns that are hard to see or help newer doctors have additional analysis on the data before making a diagnosis (CVision Lab, 2022). Overall, using AI models in the medical field shows a lot of promise to detect diseases and other health anomalies and allow patients to undergo treatment as early as possible. AI models are still improving and medical diagnosis will also improve in the years to come as complex models are better integrated.

7.2 Benefits of Instance Segmentation vs Semantic Segmentation

In this study, we used object detection to find bounding boxes of the teeth. There are two techniques: semantic segmentation and instance segmentation. On the one hand, semantic segmentation does not predict bounding boxes around the objects and treats all the different instances of the same object as the same. It finds 'teeth' and puts them all under the same level with no differentiation between teeth (even though they can be categorized by position) (Michael, 2021). On the other hand, instance segmentation can identify individual objects and differentiate each tooth (in our specific use case). As such, we have decided to use instance segmentation models in our study. These techniques have made an impact across different industries such as identifying pedestrians and lanes for self-driving cars, detecting abnormalities in MRIs and other medical scans, and mapping the world in satellite imagery (Michael, 2021).

7.3 Benefits of Segmenting Teeth vs Direct Diagnosis

TrueVa's overall goal is to be able to make better diagnosis from teeth images to help dentists but felt very strongly about making bounding boxes over predicting their first use case (overbite vs underbite vs normal teeth alignment). The reason being that with a general teeth segmentation, they have more flexibility and can explore other models that analyze the teeth specifically including applying mathematical models (previously coded by their data scientist team) which require teeth positioning as an input. Therefore, it was more beneficial to generate reliable bounding boxes thanks to our machine learning model over a classifier model for a specific use case.

8. Limitations

8.1 Dataset Characteristics

The main limitation of our model is the characteristics of the dataset we used to train it.
Firstly, the dataset size was quite limited. A well-established rule of thumb for training image detection models is that no less than 1000 images should be used per class (iMerit, 2021). Due to privacy concerns, our model had to be trained with a sample of 50 images from a public dataset. Furthermore, the 50 images we employed were taken with the aid of a cheek retractor to obtain an unobstructed view of all 16 frontal teeth. Most patients are not in possession of such a device; if TrueVA Capital wishes to implement this model as a mobile phone application for preliminary self-diagnosis, the results may consequently be suboptimal. This limitation is closely linked to the model's lower accuracy in identifying canine and bicuspid teeth compared to central and lateral incisors. The former teeth were often challenging to discern even with the naked eye, a problem which will be exacerbated by the lower-quality pictures patients will take without the cheek retractor.
Lastly, the dataset only contained pictures of relatively well-maintained adult teeth and excluded cases where patients were younger, had orthodontic appliances, missing some teeth, or had

overlapping teeth, especially where a substantial part of a tooth is concealed behind another. This meant that the model would not generalize well on test cases where there were tooth-size disparities as compared to the training data, as well as full-adult-sized teeth which were found adjacent to deciduous teeth and gaps.

9. Future Work

9.1 Improving the model

TrueVA Capital may address the dataset problem by training the model with a larger dataset that is better adapted to their needs. For the purpose of identifying malocclusion, TrueVA Capital should aim to use a minimum of 4000 randomly sampled images with similar proportions of under-, over-, cross-, and normal-bite cases. True VA could overcome the potential problem caused by the use of a cheek retractor by including in the dataset a considerable number of images taken without the device. Ideally, if this model is to be used in a mobile application, most images used to train the model should resemble those which patients would take with their mobile phones and not those taken by someone else with the help of a cheek retractor. As to the limitations arising from missing, unusually oriented, shaped, or overlapping teeth, the best way these can be tackled is by obtaining more such images to train the model with; it should also be noted that the quality of the annotations used will greatly impact the outcome of the model.

9.2 Business Opportunity: Telemedicine

Potential features
In the future, if the aforementioned limitations are overcome, the model could be used to provide patients with reliable preliminary diagnoses without assistance from any professional. The Asia Pacific telemedicine market is expected to increase from USD 10.33 billion in 2022 to USD 27.24 billion by 2027, according to the Asia Pacific Telemedicine Market Research Report. This would save time and money for both the patient and the dentist. Furthermore, the use of model could move beyond the detection of the extent of malocclusion and detect other conditions and even make suggestions and recommendations to the patient. For instance, the model could be used to suggest that the patient undergo a certain buccal treatment such as teeth realignment or teeth whitening. In the long term, the model could be used to attempt to not only suggest treatments, but also to show the patient, via descriptions and images, the degree of potential improvement, the length of time that it would take for the improvement to take effect, and the costs associated with such treatments.

Advantages

- 24/7: access to an online tool available at all hours of the day with no set working hours, increasing the convenience and accessibility to users.

- Limit spread of contagious diseases: limits doctor-patient in-person contact, meaning that were a new wave of Covid-19 or any other highly transmissible disease to rise, tele-diagnosing would continue operating.
- Accessibility to dental care: patients who are limited by time, distance, and work who do not go to the dentist as often as they should would greatly benefit from telemedicine.
- From a business perspective, telemedicine would lead to much greater efficiency. In the future, if most diagnoses were automated, dentists would be able to spend the majority of their time providing dental treatment to their patients, rather than conducting time-consuming and routinary check-ups.

Disadvantages

- As discussed, there are several difficulties associated with this project; it may take some time to implement it to the extent described above.
- A frontal picture of someone's teeth is unlikely to capture all conditions, so complete dental diagnosis automation is unlikely in the short term, and only possible in the long-term if other data, besides a frontal picture of a patient's teeth, is used.
- It is unlikely that ordinary dentist check-ups could be avoided altogether; it is hard to replace a professional's judgment with an automated decision.

10. Conclusion

The main focus of the study was to provide an accurate segmentation of teeth from dental images to aid dentists via a project led by TrueVa. We labeled ~500 images for TrueVA (from private company data) and ~50 images from an online public dataset to test our models. Once the best model was selected and built, we provided all the source codes and the labeled dataset to TrueVa so they could run the model on their internal GPUs (given the sensitivity of medical data the final trained model had to be kept separate from this project). The following is a direct quote from TrueVA's Managing Director, Swee Siong Lee, regarding our performance in this collaboration:

"The NTU team was able to help us progress with our CV development by suggesting alternative methods of approaching the modeling. The outcome was a well defined boundary of areas of interest in dental images, and provides a good basis to apply this to further develop use cases for image diagnostics. I wish to thank the team for their effort and a job well done!"

Using AI is an important asset to improve medical anomaly detection and aid healthcare professionals to diagnose patients. As such, we hope that our model can improve dental diagnosis and impact medical care thanks to deep learning.

11. References

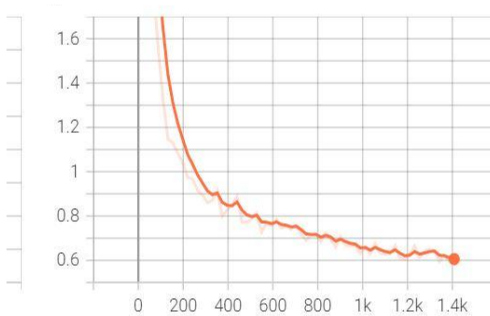
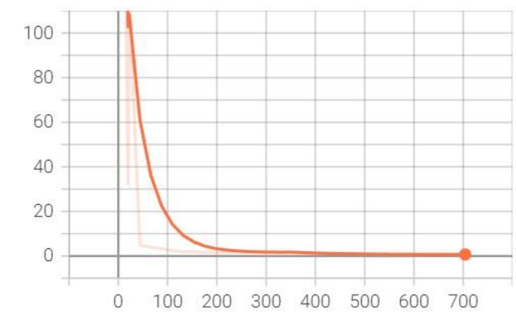
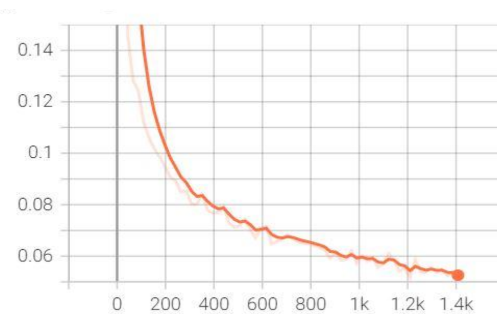
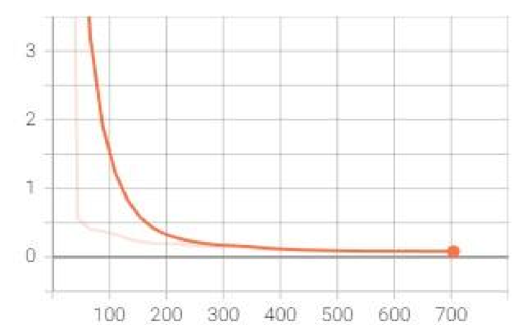
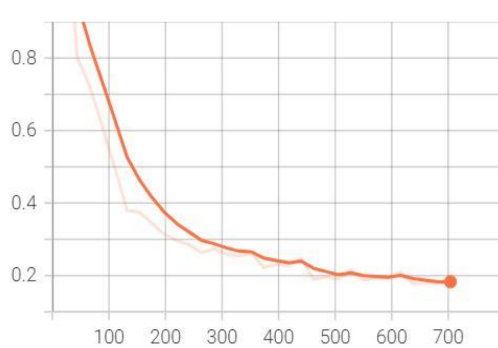
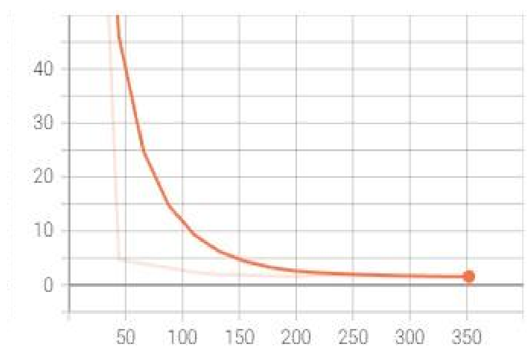
1. "Working on an AI Project? Here's How Much Data You'll Need." *iMerit*, 8 Oct. 2021, https://imerit.net/blog/how-much-data-do-you-need-for-your-ai-ml-project-all-pbm/#:~:text=Computer%20vision%20rule%20of%20thumb,1%2C000%20images%20for%20each%20class. Accessed 7 Nov. 2022.
2. Im, J., Kim, J. Y., Yu, H. S., Lee, K. J., Choi, S. H., Kim, J. H., Ahn, H. K., & Cha, J. Y. (2022). *Accuracy and efficiency of automatic tooth segmentation in digital dental models using deep learning*, 12(9429). https://doi.org/10.1038/s41598-022-13595-2
3. Michael. (2021, May 8). *Instance vs. Semantic Segmentation: What Are the Key Differences?* Keymakr. https://keymakr.com/blog/instance-vs-semantic-segmentation/
4. Tuzoff, D. V., Tuzova, L. N., Bornstein, M. M., Krasnov, A. S., Kharchenko, M. A., Nikolenko, S. I., Bednenko, G. B. (2019). *Tooth detection and numbering in panoramic radiographs using convolutional neural networks.* Dentomaxillofacial Radiology, 48(4), 20180051. doi:10.1259/dmfr.2018005
5. Shiledarbaxi, N. (2021, October 13). *Guide to solo and solov2: Ways to implement instance segmentation by location*. Analytics India Magazine. Retrieved November 7, 2022, from https://analyticsindiamag.com/guide-to-solo-and-solov2-ways-to-implement-instance-segmentation-by-location/#h-limitations-of-solo
6. *UC Santa Barbara*. (n.d.). Retrieved November 7, 2022, from https://sites.cs.ucsb.edu/~lilei/pubs/wang2020solov2.pdf
7. Wang, X., Kong, T., Shen, C., Jiang, Y., & Li, L. (2020, July 19). *Solo: Segmenting objects by locations*. arXiv.org. Retrieved November 7, 2022, from https://arxiv.org/abs/1912.04488
8. *Use mmdetection*. ArcGIS API for Python. (n.d.). Retrieved November 7, 2022, from https://developers.arcgis.com/python/guide/use-mmdetection-with-arcgis-learn/#:~:text=MMdetection%20is%20an%20open%2Dsource,the%20COCO%20object%20detection%20challenge
9. Lakshmanamoorthy, R. (2021, April 23). *Guide to mmdetection: An object detection python toolbox*. Analytics India Magazine. Retrieved November 7, 2022, from https://analyticsindiamag.com/guide-to-mmdetection-an-object-detection-python-toolbox/
10. *What is instance segmentation? [2022 guide & tutorial]*. What Is Instance Segmentation? [2022 Guide & Tutorial]. (n.d.). Retrieved November 7, 2022, from https://www.v7labs.com/blog/instance-segmentation-guide#:~:text=Instance%20Segmentation%20is%20a%20special,aerial%20crop%20monitoring%2C%20and%20more
11. *Introduction to object detection model evaluation*. (n.d.). Retrieved November 7, 2022, from https://towardsdatascience.com/introduction-to-object-detection-model-evaluation-3a789220a9bf
12. *How mask R-CNN works?* ArcGIS API for Python. (n.d.). Retrieved November 10, 2022, from https://developers.arcgis.com/python/guide/how-maskrcnn-works/
13. Khandelwal, Renu. (2019, July 31). *Computer vision: Instance segmentation with mask R-CNN*. Retrieved November 10, 2022, from

https://towardsdatascience.com/computer-vision-instance-segmentation-with-mask-r-cnn-7983502fcad1

14. Zhang, X. (2022, January 13). *Simple understanding of mask RCNN*. Medium. Retrieved November 10, 2022, from https://alittlepain833.medium.com/simple-understanding-of-mask-rcnn-134b5b330e95
15. Sharma, P. (2021, August 27). *Image segmentation Python: Implementation of mask r-cnn*. Analytics Vidhya. Retrieved November 10, 2022, from https://www.analyticsvidhya.com/blog/2019/07/computer-vision-implementing-mask-r-cnn-image-segmentation/
16. Zhou, Y. (2022, February 20). *MMDetection*. UCLA CS188 Course Projects. Retrieved November 10, 2022, from https://ucladeepvision.github.io/CS188-Projects-2022Winter/2022/02/20/team09-MMDetection.html
17. Open-Mmlab. (n.d.). *Mmdetection/customize_models.MD at master · open-mmlab/MMDETECTION*. GitHub. Retrieved November 10, 2022, from https://github.com/open-mmlab/mmdetection/blob/master/docs/en/tutorials/customize_models.md
18. (n.d.). *Tooth numbering systems*. Oral Surgeon In Cupertino. Retrieved November 10, 2022, from https://www.dentalimplantpro.com/blog/2017/tooth-numbering-systems/
19. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2018, January 24). *Mask R-CNN*. arXiv.org. Retrieved November 10, 2022, from https://arxiv.org/abs/1703.06870
20. Peter G. Peterson Foundation. (2022, February 16). Why Are Americans Paying More for Healthcare? Peter G. Peterson Foundation. Retrieved November 11, 2022, from https://www.pgpf.org/blog/2022/02/why-are-americans-paying-more-for-healthcare
21. myHSN. (2022). How doctors make a diagnosis (in 5 steps) - Health Service Navigator. myHSN. Retrieved November 11, 2022, from https://www.myhsn.co.uk/top-tip/how-do-doctors-make-a-diagnosis-5-steps
22. CVision Lab. (2022). Medical Images Analysis. CVisionLab. Retrieved November 11, 2022, from https://www.cvisionlab.com/cases/computer-vision-algorithms-for-medical-use/

11. Appendix

11.1 Visualization of Classification Loss

Number of Epochs	Classification Loss	Mask-R-CNN	SOLOv2
16	**Cross-Entropy Loss**		
	Focal Loss		
8	**Cross Entropy Loss**		

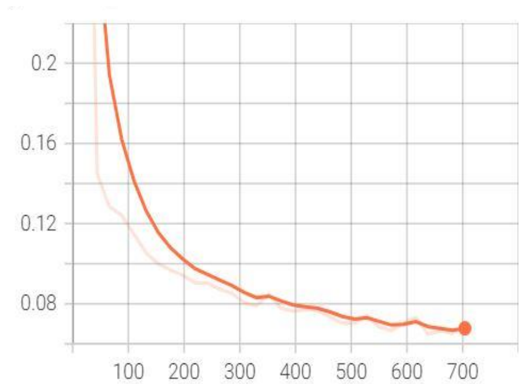
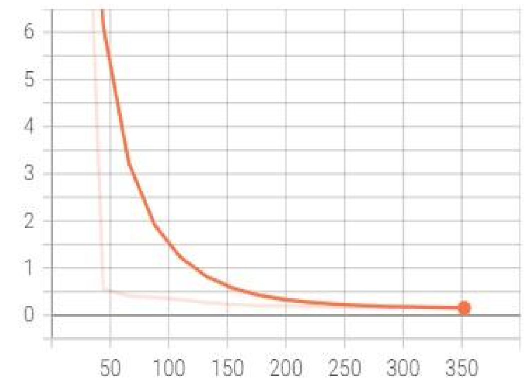
	Focal Loss		

Table 6: Visualization of Classification Loss