DentalVision: Automated Teeth Segmentation for Dental Diagnostics using Vision Transformers

Abhishek Kumar Singh (M23CSA503)
Rishabh Sharma (M23CSA532)

April 4, 2025

Contents

1 Objective

The objective of this project is to develop an automated tool that segments teeth in dental X-ray images using deep learning techniques. The tool should provide accurate visualizations and be accessible through a user-friendly web application to assist dentists in diagnostics and treatment planning.

2 Approaches Used

- **Model**: Vision Transformer (ViT) as the encoder and a custom lightweight U-Net-inspired decoder.
- **Data Preprocessing**: Images and masks resized to 224x224; grayscale normalization; binary thresholding.
- **Training**: Binary Cross Entropy loss with Adam optimizer, trained using PyTorch and Hugging Face's 'transformers' library.
- **Interface**: Developed an interactive web app using **Streamlit**.

3 Dataset

- **Source**: Kaggle - Teeth Segmentation Dataset
- **Content**: Panoramic dental X-rays and binary masks labeled by humans.
- **Format**: Separate folders for images and masks, available in PNG and JSON format.

4 Model Architecture

We use a **Vision Transformer (ViT)** as an encoder and a custom decoder (inspired by U-Net) to perform segmentation.

- **Encoder**: ViT with patch size 16, hidden size 768, and 6 transformer layers.
- **Decoder**: A stack of Conv2D, ReLU, and ConvTranspose2D layers for upsampling.
- **Output**: 1-channel sigmoid-activated mask for binary segmentation.

5 Training and Evaluation

- **Loss Function**: Binary Cross Entropy
- **Optimizer**: Adam (learning rate = 0.0001)
- **Batch Size**: 4
- **Epochs**: 5

- **Train-Validation Split**: 80% train, 20% validation
- **IoU**: 0.82
- **Dice Coefficient**: 0.89

6 Results

- The model achieved strong performance on the test set:
  - IoU (Intersection over Union): **0.82**
  - Dice Coefficient: **0.89**
- The model generalizes well to unseen X-rays and overlays masks accurately.

7 Deployment Link

The model is deployed and accessible through a live Streamlit app:
Live App: `https://dentalvision-vitonet.streamlit.app`

8 GitHub Repository

You can find the complete source code and training scripts on GitHub:
Repository URL: `https://github.com/akshziitj/dentalvision-vitonet`

9 Screenshots from Deployed Website

Figure 1: Upload interface for dental X-rays

Figure 2: Results: Original image, mask prediction, and overlay

10 Model Saving and Inference

- Trained model saved as `vit_teeth_segmentation.pth`
- Easily loaded for predictions

```
model.load_state_dict(torch.load('vit_teeth_segmentation.pth'))
```

11 Project File Structure

```
dentalvision-vitonet/
|-- model/                              # Directory containing trained models
|   |-- vit_teeth_segmentation.pth      # Trained ViT segmentation model weights
|-- sample_data/                        # Sample inputs and outputs
|   |-- sample_image_01.jpg             # Example input X-ray image
```

```
|   |-- sample_result.jpg                # Corresponding predicted mask
|-- .gitattributes                      # Git LFS and attributes configuration
|-- .gitignore                          # Files and folders to ignore in git
|-- DentalVision_CV_Project.ipynb       # Jupyter notebook for training pipeline
|-- README.md                           # Project overview and instructions
|-- Report.pdf                          # Final project report
|-- requirements.txt                    # Python dependencies list
|-- streamlit_app.py                    # Streamlit-based web application
|-- utils.py                            # Helper functions: preprocessing, metrics, etc
```

12 Conclusion

DentalVision successfully demonstrates how Vision Transformers can be leveraged for medical image segmentation tasks. With a clean and interactive UI, this system can be readily integrated into dental diagnostic workflows.

13 Future Work

- Integrate instance segmentation to identify individual teeth
- Enable detection of anomalies (cavities, missing teeth, etc.)
- Deploy a REST API version for clinical integration
- Enhance model with Swin Transformers or SAM

14 Acknowledgments

- Kaggle: For the open-source dental X-ray dataset
- Hugging Face: For the Vision Transformer model base
- Streamlit: For the lightweight UI framework

15 References

1. Dosovitskiy, A. et al. (2021). *An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale*. ICLR. https://arxiv.org/abs/2010.11929
2. Ronneberger, O. et al. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. MICCAI. https://arxiv.org/abs/1505.04597
3. Streamlit Documentation: https://docs.streamlit.io
4. PyTorch Documentation: https://pytorch.org
5. Hugging Face Transformers: https://huggingface.co/docs/transformers/index
6. Kaggle Dataset: Teeth Segmentation Dataset on Kaggle