

# My Byte

## Requirements

- Fitment Prediction of tooth aligner
  - Fit Type
    - Best fit
    - Moderate fit
    - Fair fit
    - Poor fit
  - Yolo v8 Image Segmentation & Classification
- Frame Picker for the fitment prediction model
  - This model picks the best frame from user via **Camera** for the segmentation model

## POC for Frame Picker

- MediaPipe
  - Pose Detection
    - Detects when the image contains full facial region.
    - Fails to detect when the eye region or nose is unavailable.
  - MobileNet
    - Freezed Accuracy 30%
    - Unfreezed Accuracy 40%
  - EfficientNet
    - Freezed Accuracy 35%
    - Unfreezed Accuracy 50%
- EfficientNet xl
  - Freezed Accuracy 50%
  - Unfreezed Accuracy 65%

## Dataset Preprocessing

Classification Image Data of 30000 with 6 classes which has people's face as mouth opened.

- *Left*
  - *Right*
  - *Upper*
  - *Lower*
  - *Smile*
  - *Discomfortable*
1. While observing the **Discomfortable** category the images are in irregular position and so it is omitted.
  2. Left and Right classes are shuffled and need to be isolated.
    - a. In order to isolate EfficientNetXL is used.
    - b. Manually sorted 100 images from each classes are taken from shuffled images.
    - c. Trained the model with 90% of Accuracy.
    - d. Predicted the remaining shuffled images to further classification.
    - e. Thus Left and Right images are isolated.
  3. Dataset is configured to Balanced.

## Training

- Training data 70% | Testing data 15% | Validation data 15%
- Input Image shape (768, 768)
- EfficientNetXL provides 65% of Validation Accuracy initially.
- After removing the discomfort class model gives 75% of Validation Accuracy.
- Isolating the Left and Right class model gives 88% of Validation Accuracy.
- Further configuring the dataset to Balanced, model gives 96.7% of Accuracy.

## Model Compression

- Size of the EfficientNetXL model is 1.5 GB.
- Moved to **EfficientNetv2B0** model is 25 MB, Accuracy remains 96.7%.
- After **Quantization** model reduced to 6.5 MB, Accuracy remains 96.5%.
- To reduce model size further moved to custom model development.
  - 8 Convolution layers followed by feed forward.
  - Iteration
    1. Accuracy 75%.
    2. After several attempts of *remodifying the layer pattern* & obtained 85% of Accuracy.
    3. After several attempts of *Tuning the hyper-parameter* model gives 96% of Accuracy & model size is 2 MB.
    4. Further quantising model reduced to 200 KB & gives 95.6% of Accuracy.
  - TFLite is used for quantization.
- Reduced the input shape from 768 \* 768 to 224 \* 224 improves the inference speed from 10 to 60 fps & the accuracy remains 95.6%.
- Further reduced the shape to 64 \* 64 increases 220 fps & accuracy remains 95.6%.
- This model is deployed in Android & ios using ReactNative.

## Frame Picker Algorithm

- This guide the user to select 5 best frames from each category.
- Initially the UI guide the user to focus the left side of the face.
- Variance of Laplacian is applied on Live Camera frames which gives Laplacian score.
- Higher the score, more the quality of frame.
- Top 30 frames with highest Laplacian score is passed to the model.
- Best frame with highest confidence score is selected.
- This frame is fed to the fitment prediction model and segmentation & classification is carried on.
- Yolo model returns segmentation coordinates as well as the prediction class.