MULTI-MODAL EMOTION RECOGNITION

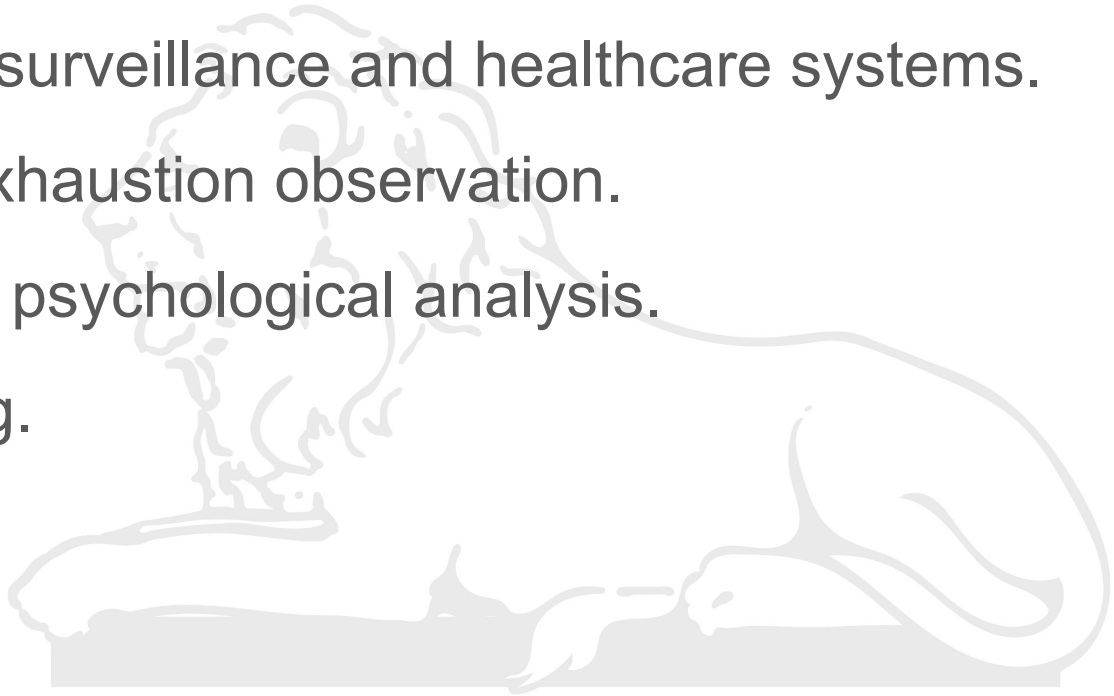
BAVESH BALAJI
COE18B007

INTERNAL GUIDE: Dr. V. MASILAMANI
EXTERNAL GUIDE: Dr. PARTHA PRATIM ROY

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY,
DESIGN AND MANUFACTURING,
KANCHEEPURAM

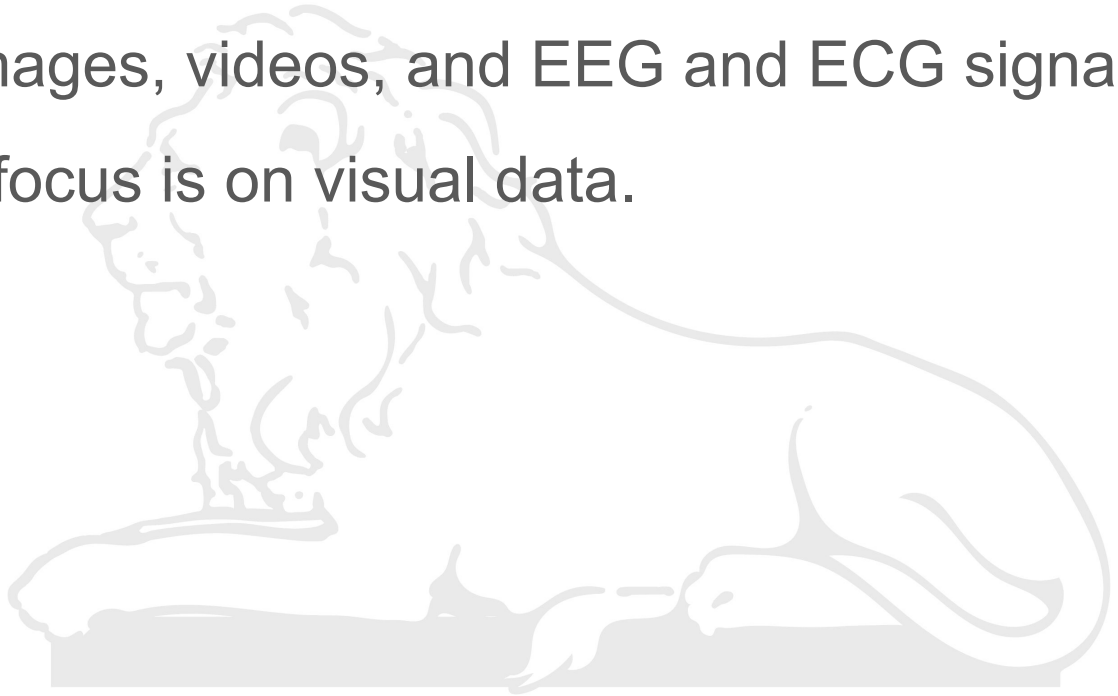
MOTIVATION

- Helps in understanding the psychological state of mind of people.
- Used in surveillance and healthcare systems.
- Driver exhaustion observation.
- Criminal psychological analysis.
- Teaching.



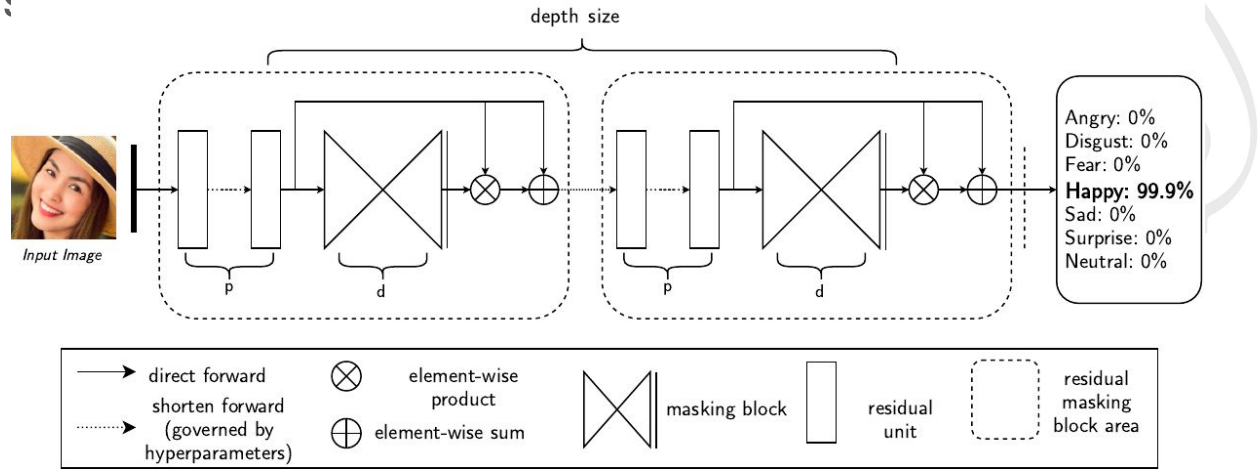
PROBLEM STATEMENT

- Performing emotion recognition using different modes of data.
- Using Images, videos, and EEG and ECG signals.
- Current focus is on visual data.



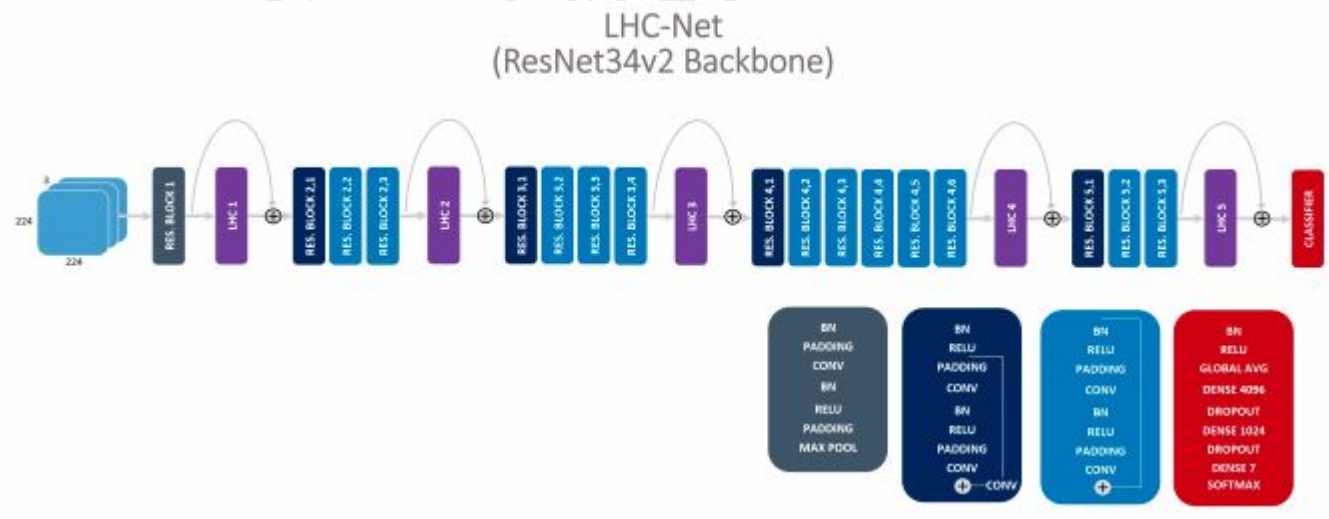
LITERATURE SURVEY

- CNN based models like VGG, Resnet.
- Residual Masking Network:-
  - Uses residual blocks from ResNet to encode the features of the images.
  - Uses U-Net based encoder-decoder model after every residual block.
  - gives attention to more important features



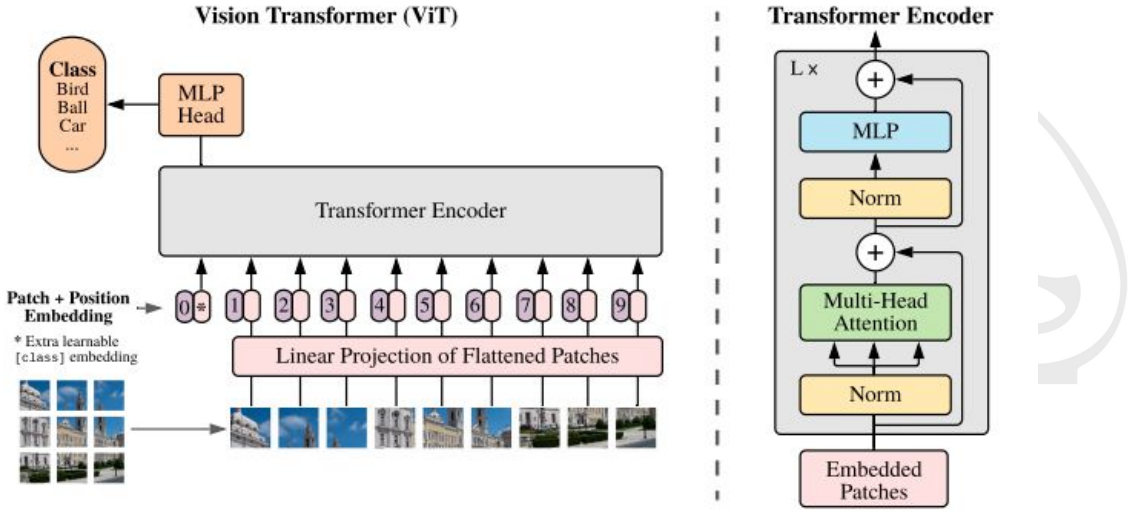
LITERATURE SURVEY

- Residual Attention Network:-
  - Only attention-based model.
  - uses self-attention on the features extracted from them image using a CNN.



VISION TRANSFORMER

- input image split into different square patches and embedded.
- encoder uses self-attention to find out the relationship between different patches.



WHY VISION TRANSFORMER

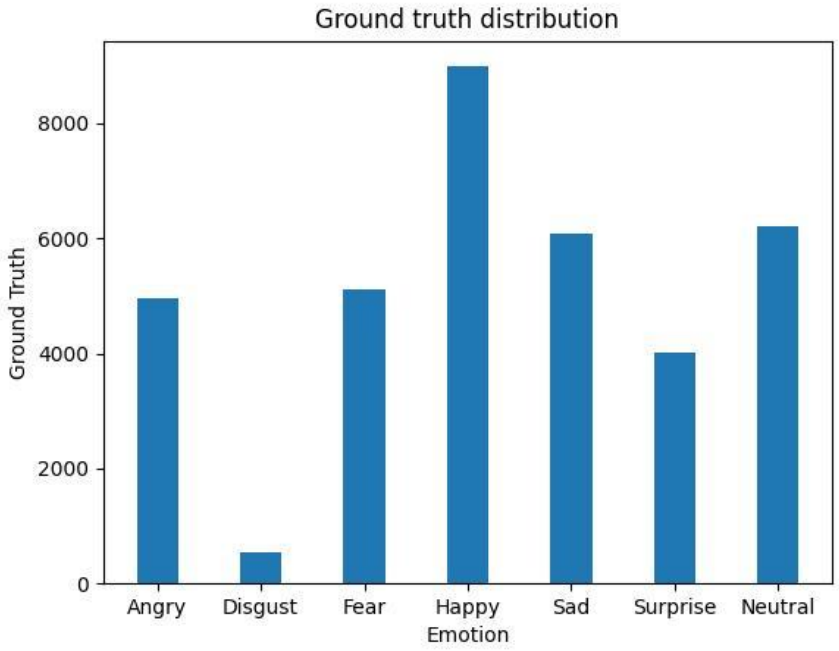
- Existing models are all mostly CNN based.
- All major models use attention on the features directly or indirectly.
- No model uses self-attention on the input image.

WORK DONE

- Data:-
  - FER013 dataset.
  - 35k images with 7 basic emotions.
  - grayscale images of size (48, 48).

WORK DONE

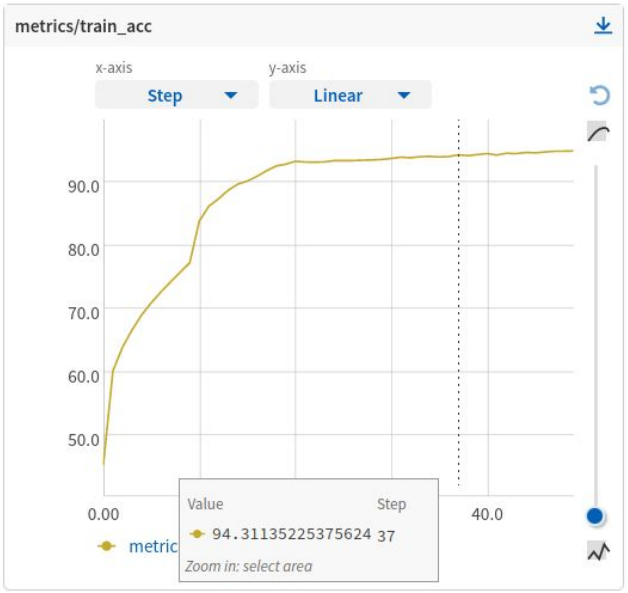
- Data:-
  - Highly class imbalanced.

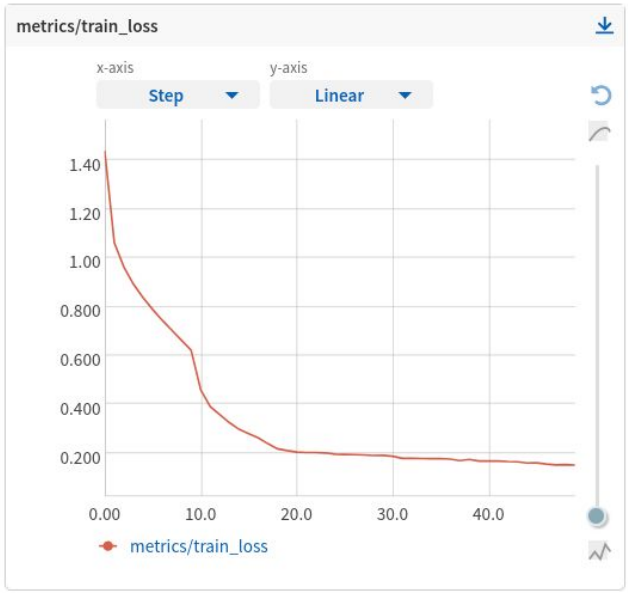


WORK DONE

IMPLEMENTING RESIDUAL MASKING NETWORK

- To understand the functioning of the masking block and the data in hand.
- Results:-

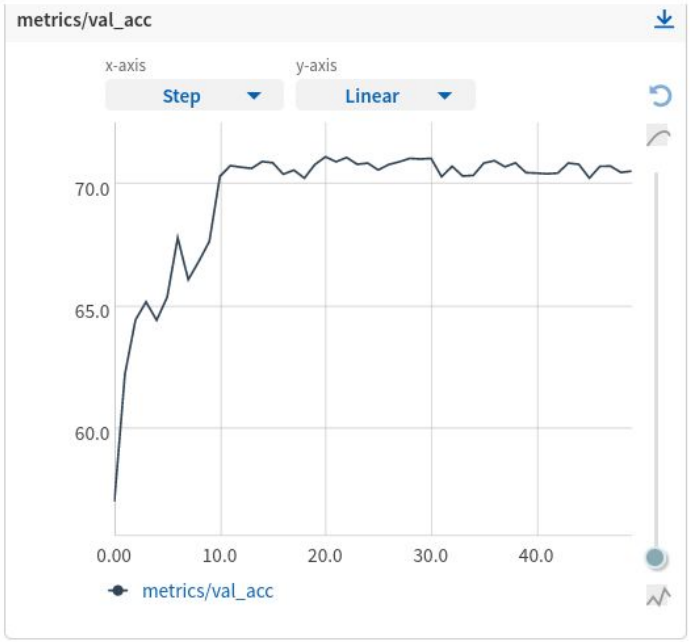


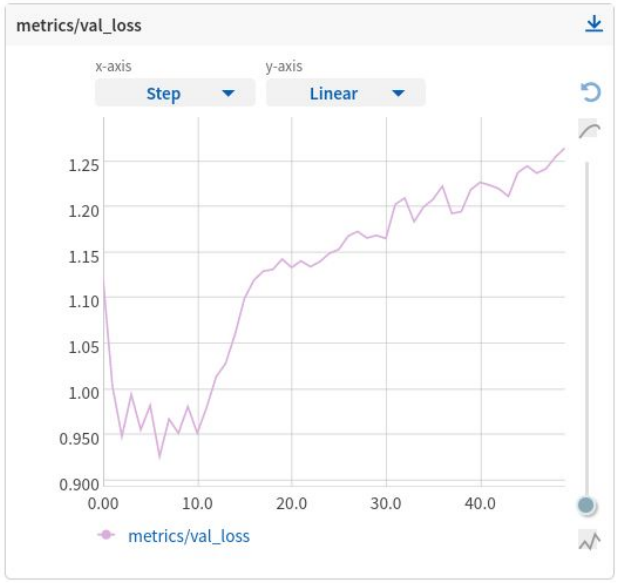


WORK DONE

RESIDUAL MASKING NETWORK

- Results:-





WORK DONE

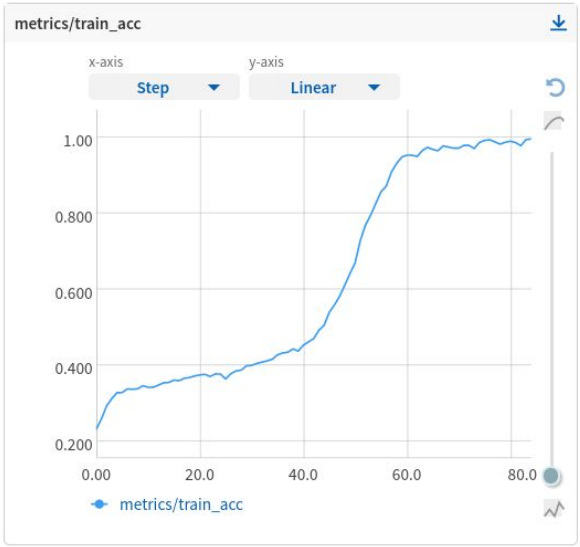
VISION TRANSFORMER:-

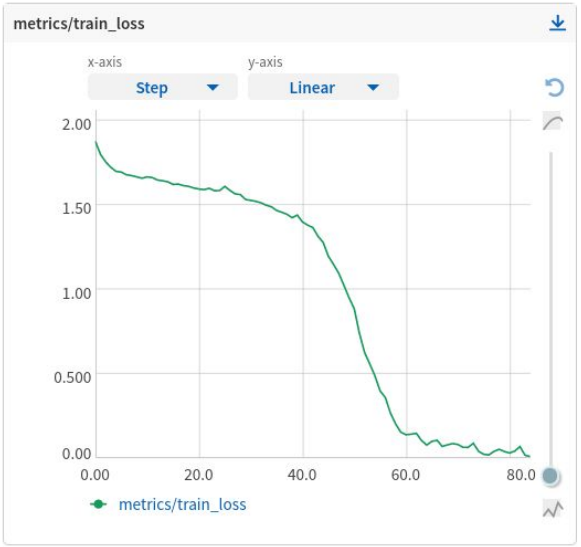
- Experimented various data augmentations, initializations and optimizers.
- RandomHorizontialFlip, RandomRotation and random erasing worked best.
- SGD converged faster, but Adam gave us better performance.
- Used truncated normal distribution to initialize weights.

WORK DONE

VISION TRANSFORMER:-

- Results:-
  - Used PyTorch and the timm library for training.
  - MLOps tools such as neptune and hydra for experiment tracking.

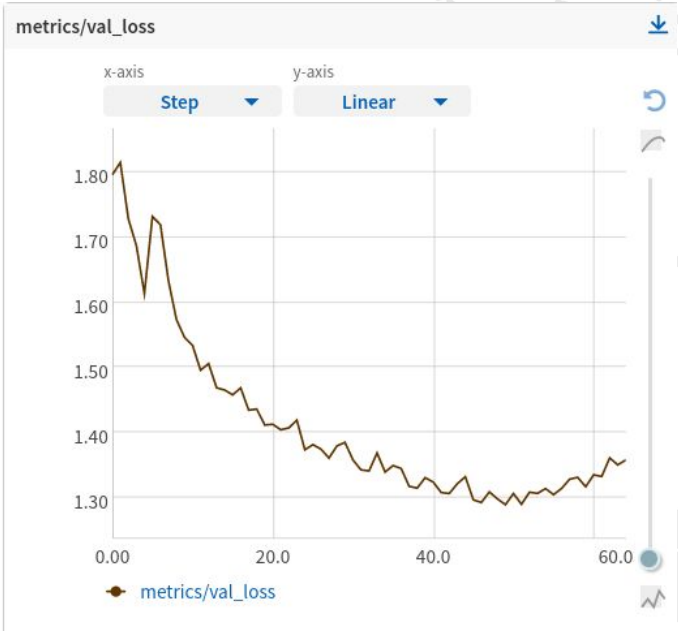


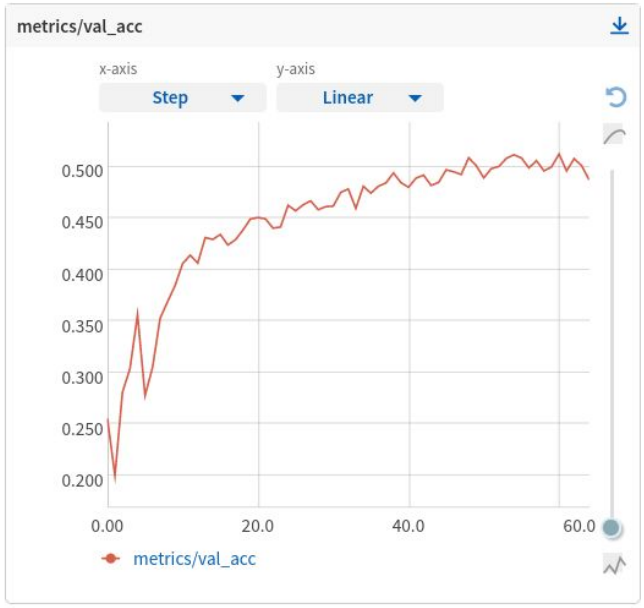


WORK DONE

VISION TRANSFORMER

- Validation results:-





WORK DONE

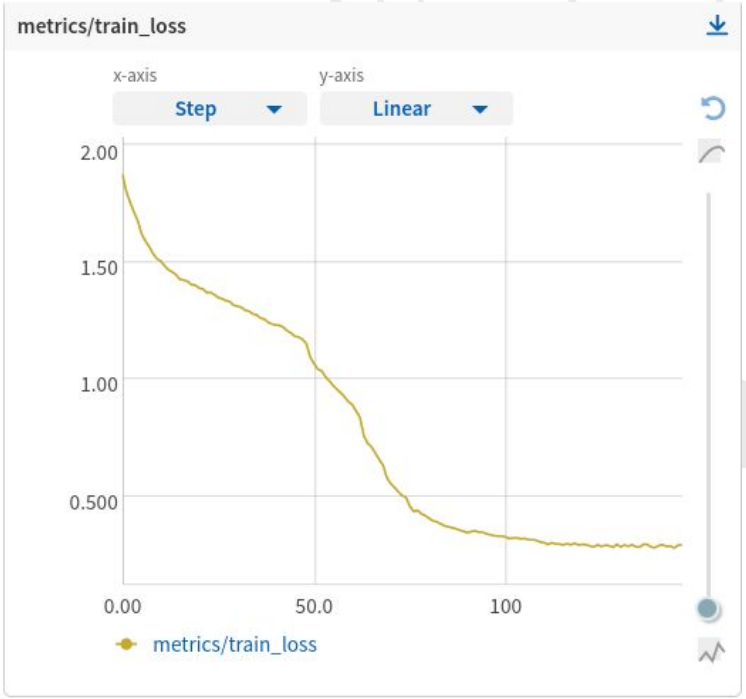
DATA EFFICIENT IMAGE TRANSFORMER

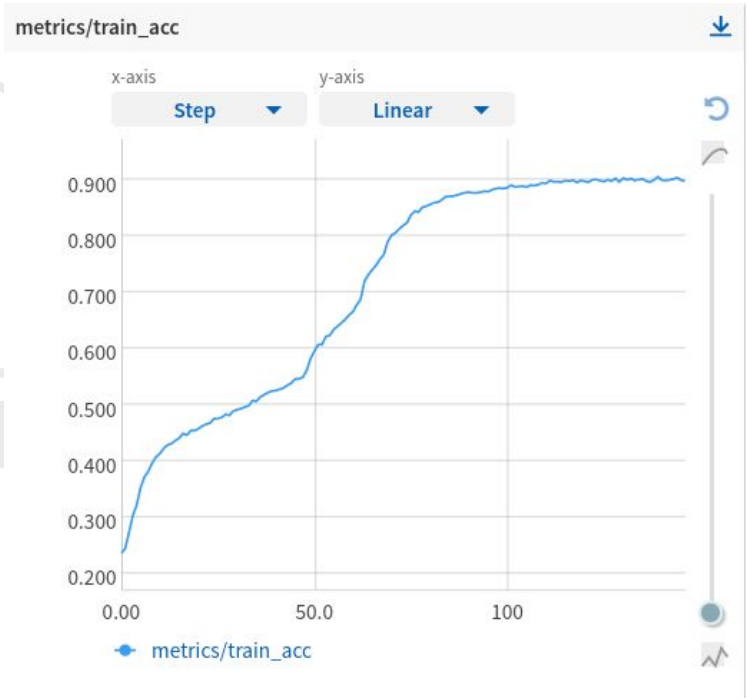
- Vision transformers did not give us desired result(52% val accuracy and 54% test accuracy).
- One reason:- insufficient amounts of data.
- Decided to use DeiT, a modified version of ViT.
- Optimized to work well for small data.

WORK DONE

DATA EFFICIENT IMAGE TRANSFORMER

- Results:-
  - All training strategies used - same as ViT except Stochastic depth.
  - Optimizer - AdamW.

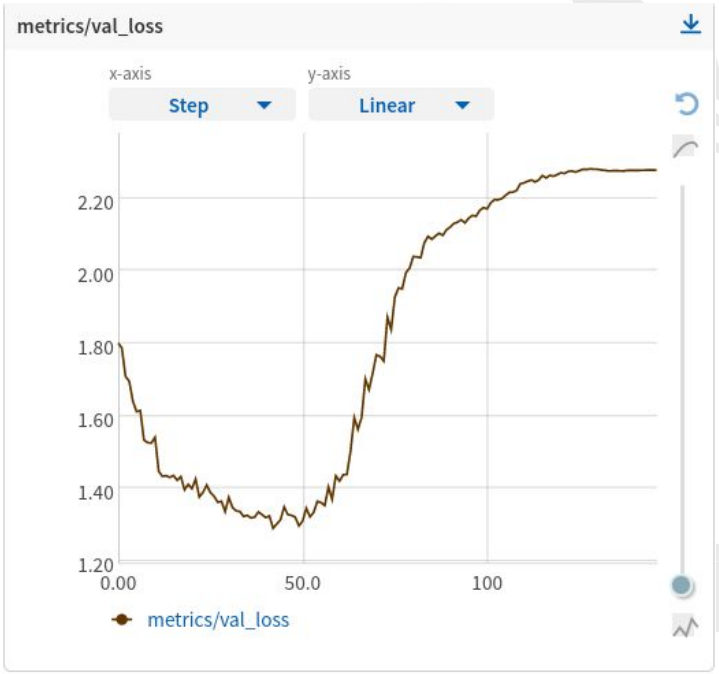


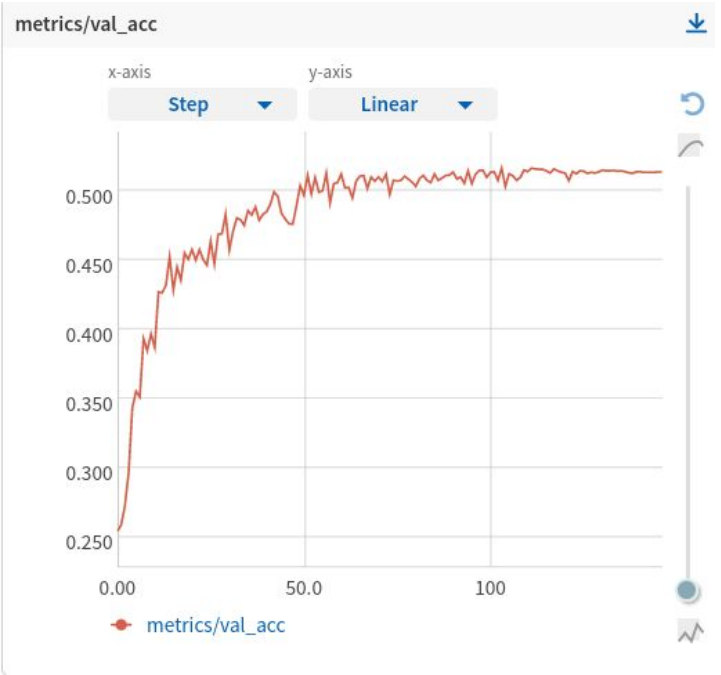


WORK DONE

DATA EFFICIENT IMAGE TRANSFORMER

- Validation Results:-

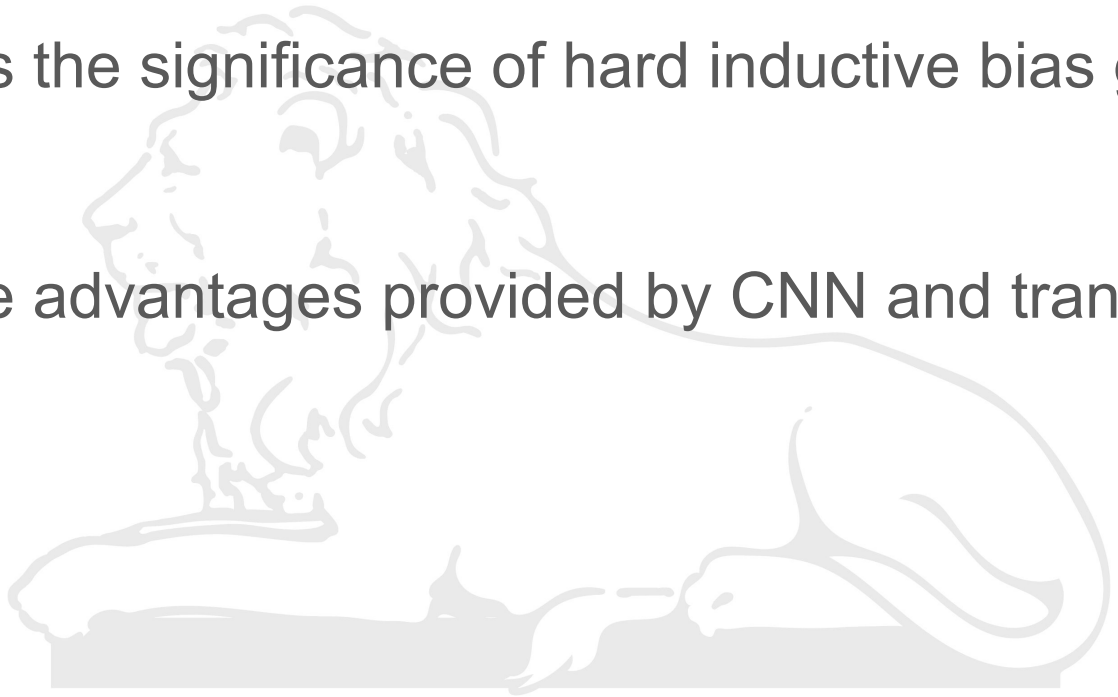




WORK DONE

CNNs vs TRANSFORMERS

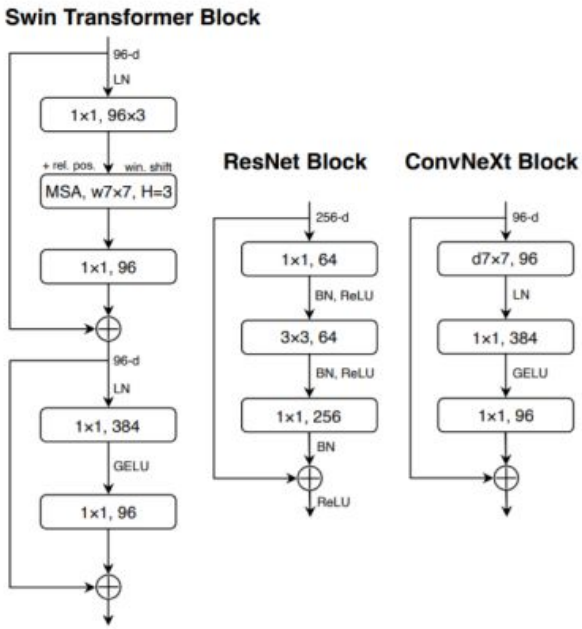
- ViT and DeiT - poor performance.
- Indicates the significance of hard inductive bias given by CNN.
- Combine advantages provided by CNN and transformers.



WORK DONE

ConvNeXts

- Fully convolutional architecture.
- Incorporates key components from hierarchical transformers.
  - ‘Patchifying’ stem using non-overlapped convolution layers.
  - Using depthwise convolutions.
  - Using GELU instead of ReLU



WORK DONE

ConvNeXts

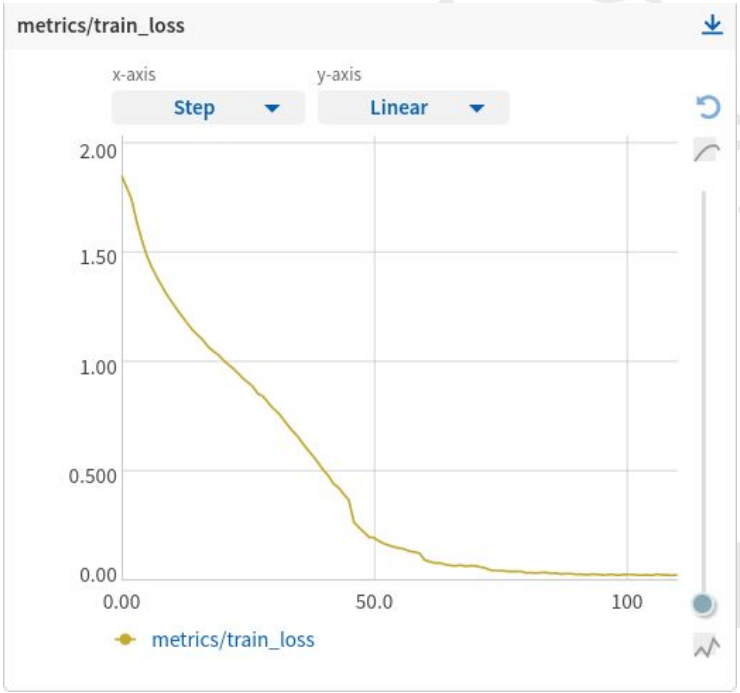
Training Details

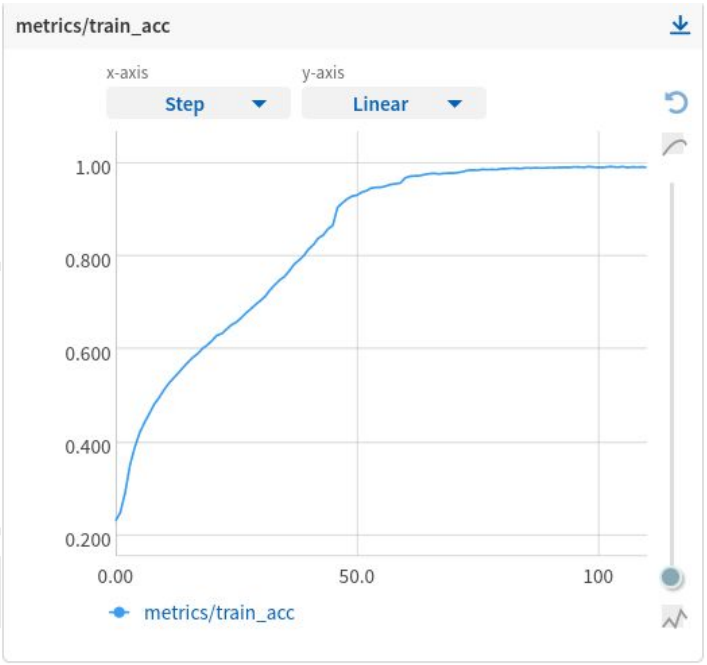
- Used random augment, CutMix and Random erasing.
- Initialization - Truncated Normal distribution.
- Stochastic Depth and Layer Norm for regularization.
- Optimizer - AdamW.

WORK DONE

ConvNeXts

- Training Results:-

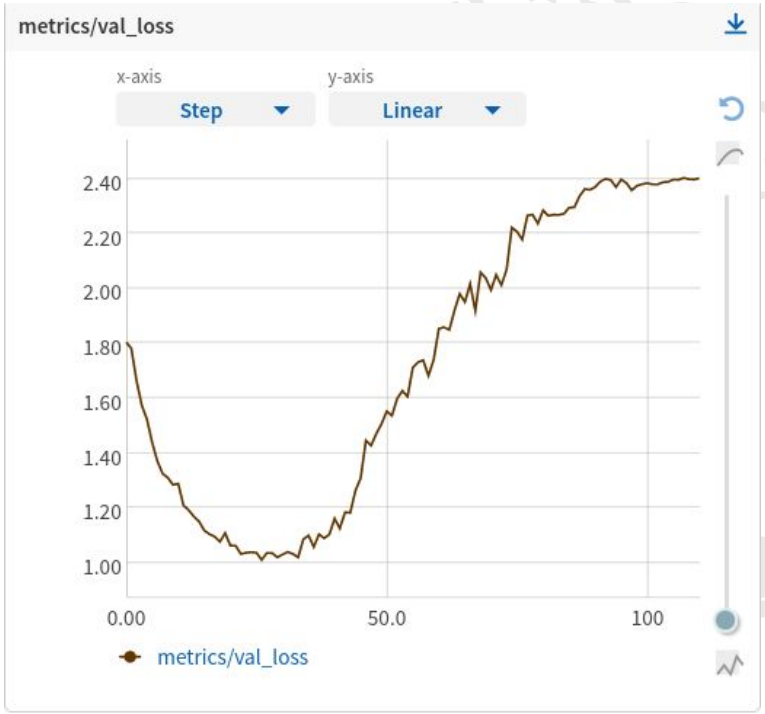


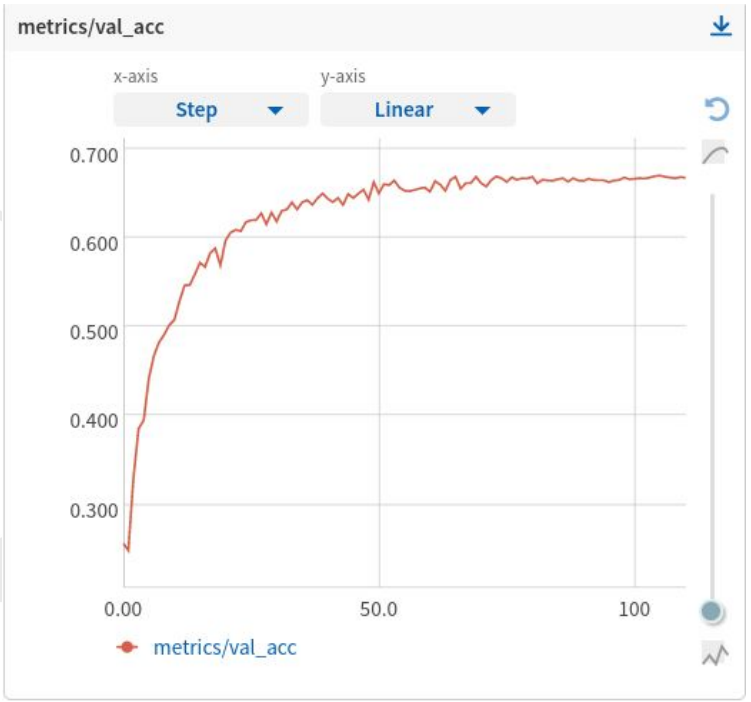


WORK DONE

ConvNeXts

- Validation Results:-

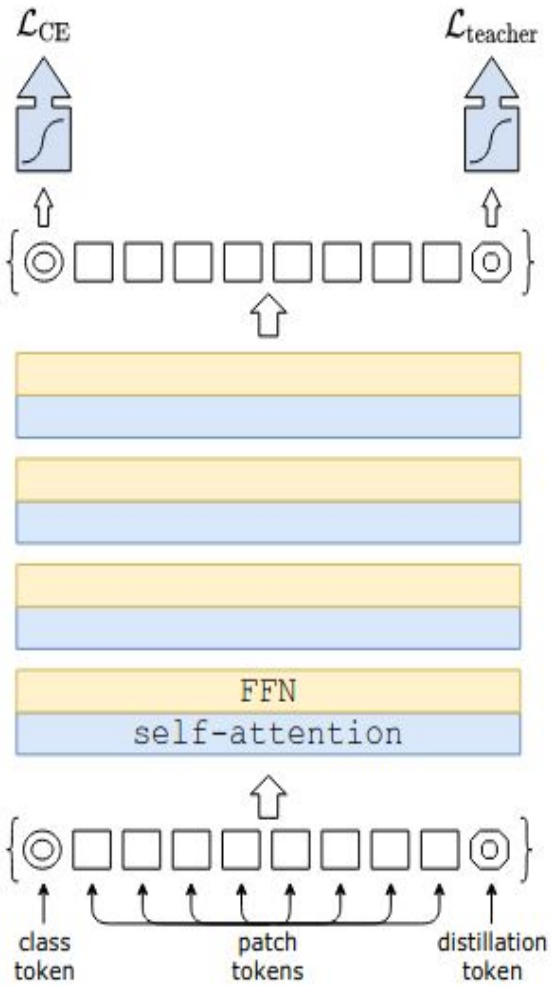




WORK DONE

Repeated Knowledge Distillation:-

- Teacher-student strategy.
- Student learns from teacher using self-attention.
- Use this strategy to train a student model which acts as a teacher to train another student model.



EEG EMOTION RECOGNITION

EEG:-

- Records the electrical activity of brain.
- Fluctuations indicate sudden change in brain activity.
- Evoked potentials:- averaged out signals over multiple epochs.
- Used for diagnostic study such as epilepsy, sleep disorder.

WORK DONE

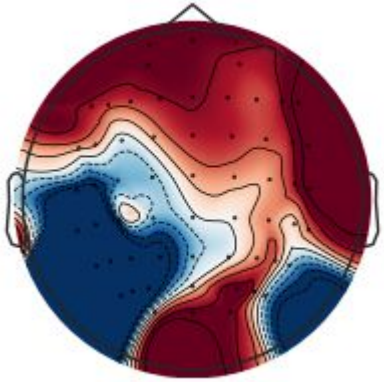
Data pre-processing:-

- Rejecting Noise from raw data.
  - Downsampling data.
  - Re-referencing data
  - fourier transform(time -> frequency domain)
  - Apply bandpass filter
  - Independent component analysis.
- pre-processed data -> 2-dimensional images using topography.
- Used MNE for all the operations.

WORK DONE

Data:-

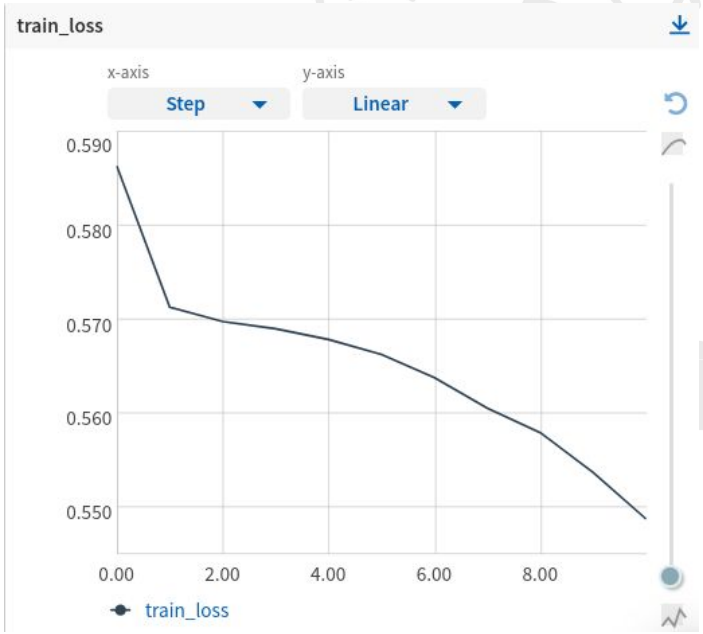
- Private dataset which is collected by ourselves of 122 subjects.
- Consists of 1300 topographical maps of majorly 2 categories:- happy and not happy.



WORK DONE

Training and results:-

- Used basic ResNet models to create a baseline.
- No data augmentations.
- Basic regularization and initialization techniques.

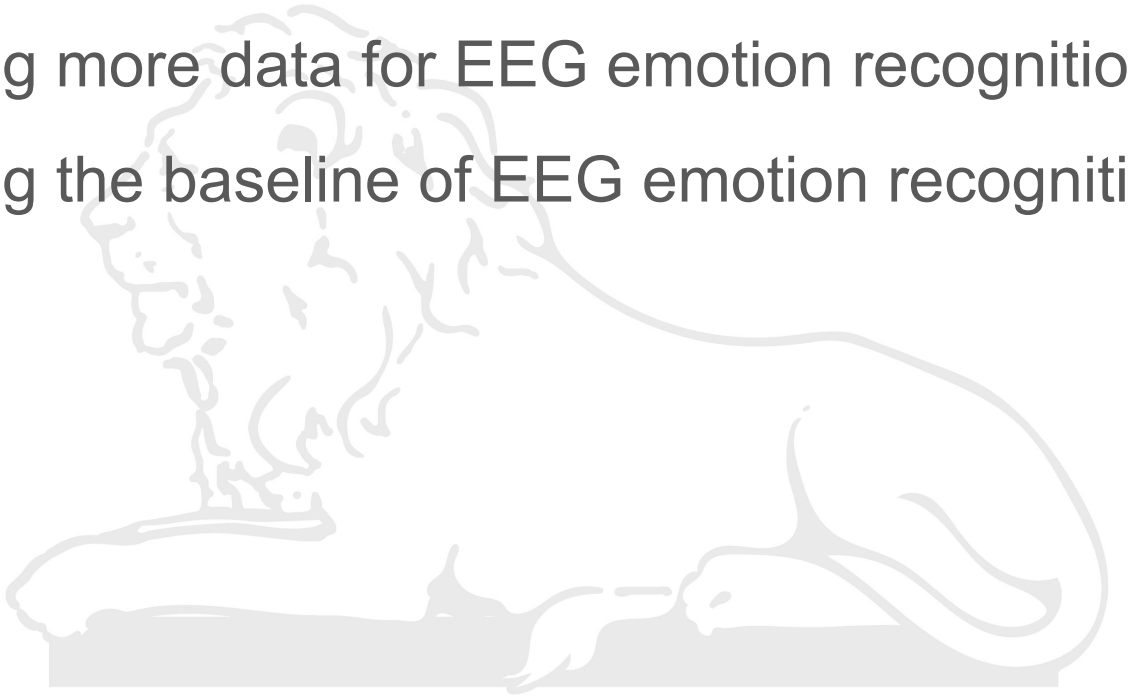


WORK DONE

Validation Results:-

PLAN OF ACTION

- Complete implementing the knowledge distillation strategy and test it out on our dataset.
- Collecting more data for EEG emotion recognition.
- Improving the baseline of EEG emotion recognition.



REFERENCES

- ResMasking Network:- https://ieeexplore.ieee.org/document/9411919
- Residual Attention Network:- https://arxiv.org/pdf/2111.07224v2.pdf
- An Image is worth 16x16 words:- https://arxiv.org/pdf/2010.11929.pdf
- Attention is all you need:- https://arxiv.org/pdf/1706.03762.pdf
- FER2013 Dataset:- https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/data

Thank You

Open to
Questions