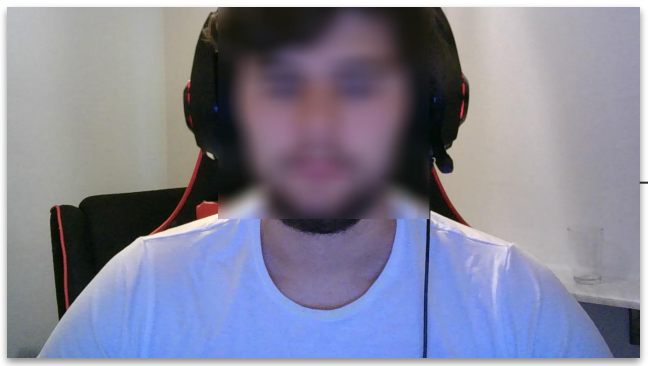
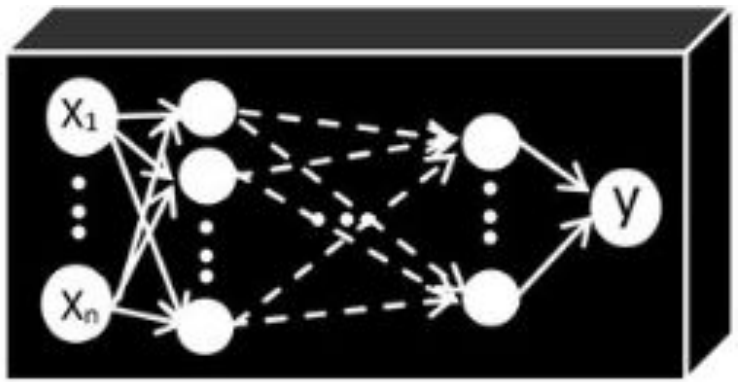
Working posture analysis

CNNs for classifying working posture using a simple webcam

By Gustav Tindberg & Victor Schack, Supervised by Marcus Klang

The problem





GoodPosture

BadPosture

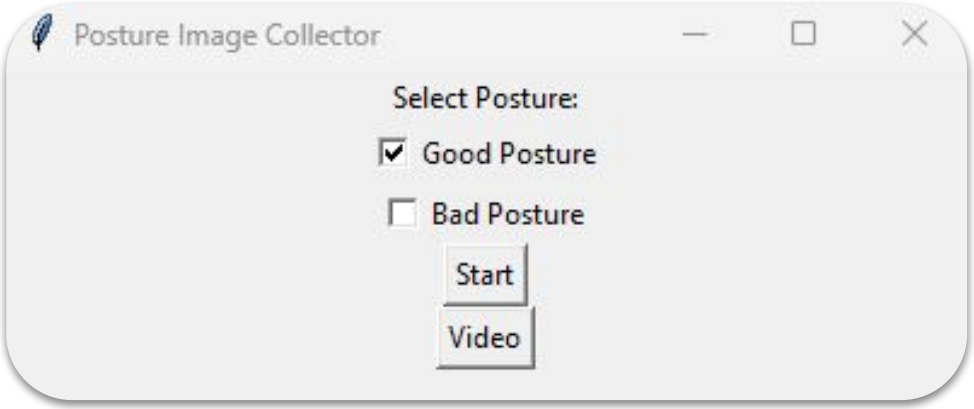
The problem

- Poor working posture in the office
- Only marginal overlap with prior academic work
  - Multiple angles
  - Wearables
  - Setup procedures before each use
- Limited to common hardware
- No suitable existing data available.
  - Collect new dataset from scratch.

Data collection

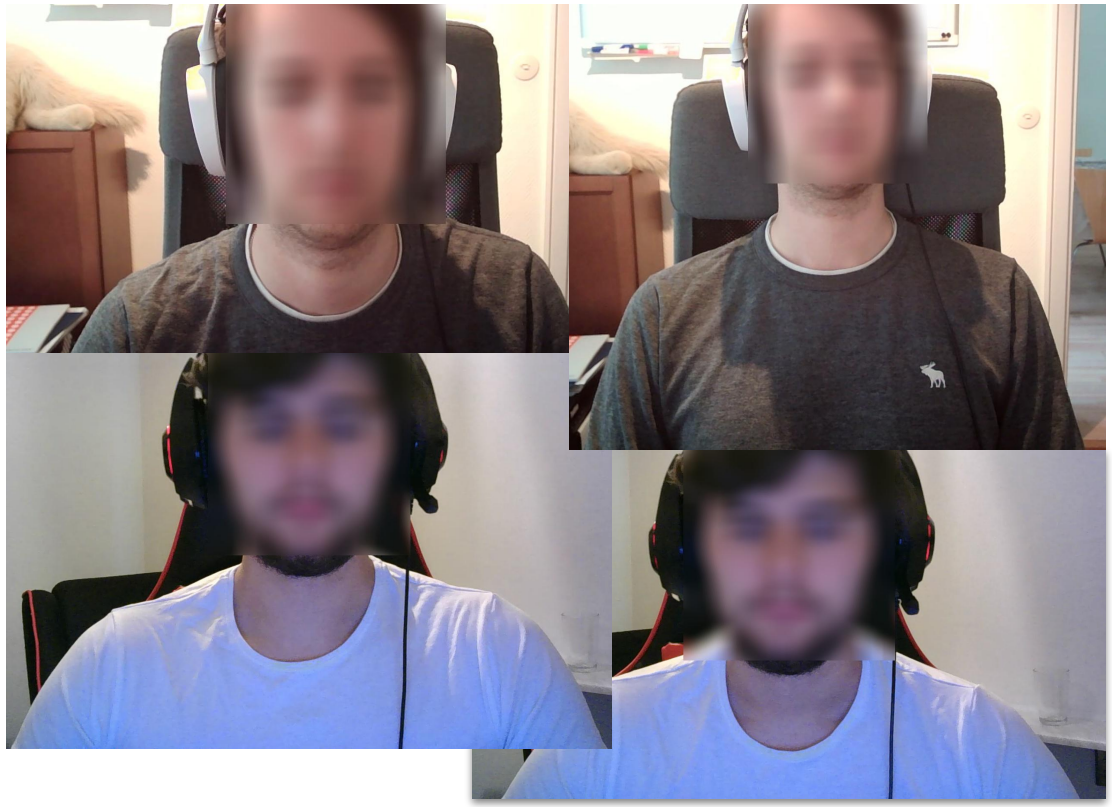
Data collection

- Live collection
- Video collection



Data structure

- To not mix subjects in splits
- Modular datasets



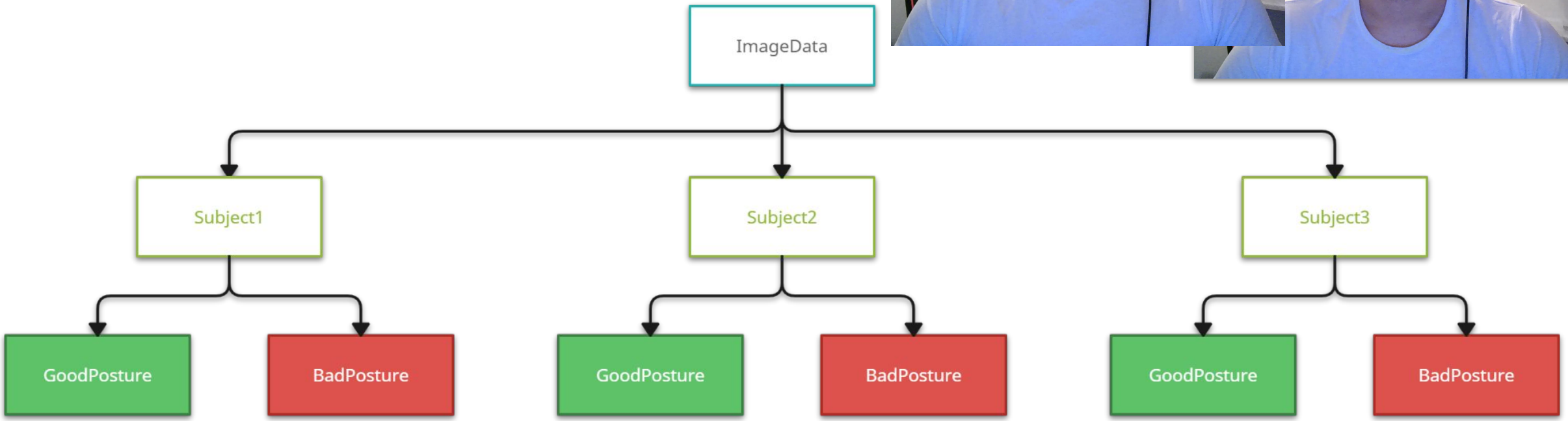


Image Augmentations

- First added Transforms.
  - Helped with generalization
  - Mitigating overfitting

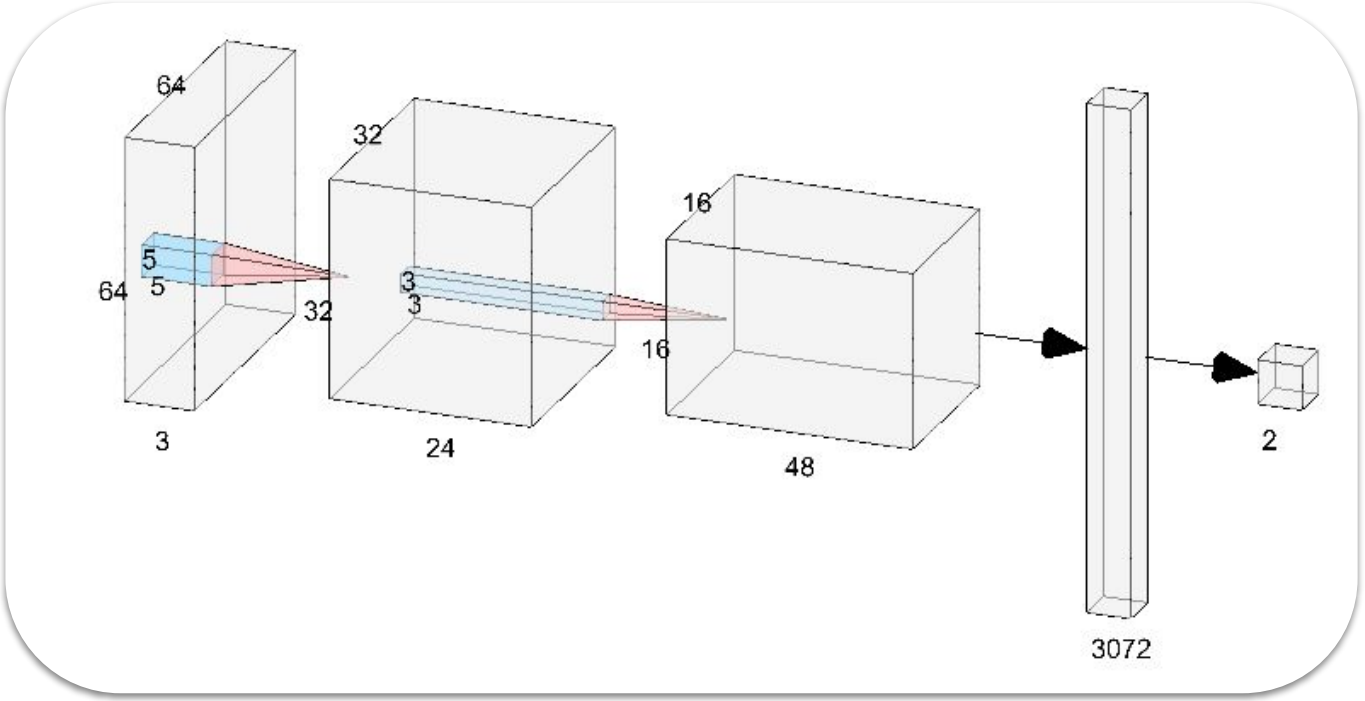
```
transform = transforms.Compose([
    transforms.RandomPerspective(0.1,p=0.3),
    transforms.RandomResizedCrop(size=size),
    transforms.TrivialAugmentWide(),
    transforms.ColorJitter(brightness=0.2, contrast=0.2, saturation=0.2, hue=0.0),
    MaxVit_T_Weights.IMAGENET1K_V1.transforms(),
    ])
```

- Cropping based on head.
  - Centering the subject
  - Reducing background

Models

Custom CNN

- Simple Architecture.
- Trained from scratch.
- Specialized to our task



Transfer learning

- Pre-trained and optimized
- Great at solving feature extraction
- Performance out of the box
- Evaluated on ImageNet
    - Object detection task
- Working posture detection
    - Aesthetic assessment
    - Pose estimation

Transfer learning

- RegNet (_Y_3_2GF)
  - Good performance at lower network scales
  - Low amount of parameters and compute
  - Reliable starting point
- MaxVit
  - Recently added to Torchvision
  - Also visual aesthetic assessment
  - Only slight increase in Params from Regnet

Weight	Acc@1	Acc@5	Params	GFLOPS
MaxVit_T_Weights.IMAGENET1K_V1	83.7	96.722	30.9M	5.56
RegNet_Y_3_2GF_Weights.IMAGENET1K_V2	81.982	95.972	19.4M	3.18

Hyper-parameter tuning

Hyper-parameter tuning

Using HyperOpt on the MaxVit network

- Starting point
  - lr: 0.02, wd: 1e-4, epochs: 40
- First attempt:
  - 15 Trials, 1 hour runtime
  - lr: 0.001, wd:1e-5, epochs: 80
- **Final attempt:**
  - 50 Trials, 8 hour runtime
  - Parameter space:
    - lr: 1e-5 to 1e-3
    - wd 1e-10 to 1e-4
    - epochs: [60,70,80,90,100]
  - **lr: 0.0005, wd: 1e-4, epochs: 70**

Tuned by minimizing loss

Evaluation

LOGO, Cross-validation

- Stratified on Subjects
  - Training one model per Subject
  - Leaving that Subject out for testing
- Used to evaluate a generalized performance of the networks
  - Looking at the performance on each subject
  - Looking at the mean across all models

MCC: Matthew's correlation coefficient

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP}+\mathrm{FP})(\mathrm{TP}+\mathrm{FN})(\mathrm{TN}+\mathrm{FP})(\mathrm{TN}+\mathrm{FN})}}$$

Results

LOGO Cross-validation results

On all available data

- CNN
  - Combined F1: 0.627
  - Mean MCC: 0.250
- RegNet
  - Combined F1: 0.57
  - Mean MCC: 0.173
- MaxVit (Hyper-parameter tuned using HyperOpt)
  - Combined F1: **0.633**
  - Mean MCC: **0.343**

MCC scores

MaxVit

```
Ali  mcc: -0.2948839123097942 7
Ami  mcc: -0.0944911182523068 1
Ebb  mcc: 0.4166666666666667
Eri  mcc: -0.2212488394343549
Fre  mcc: 0.832502943378437
Gus  mcc: 0.6546536707079772
Jac  mcc: 0.919866211007799 9
Jes  mcc: 0.5144957554275265
Lou  mcc: 0.7337993857053428
Mar  mcc: 0.1809068067466581 6
Vic  mcc: 0.1354322446219716 2
```

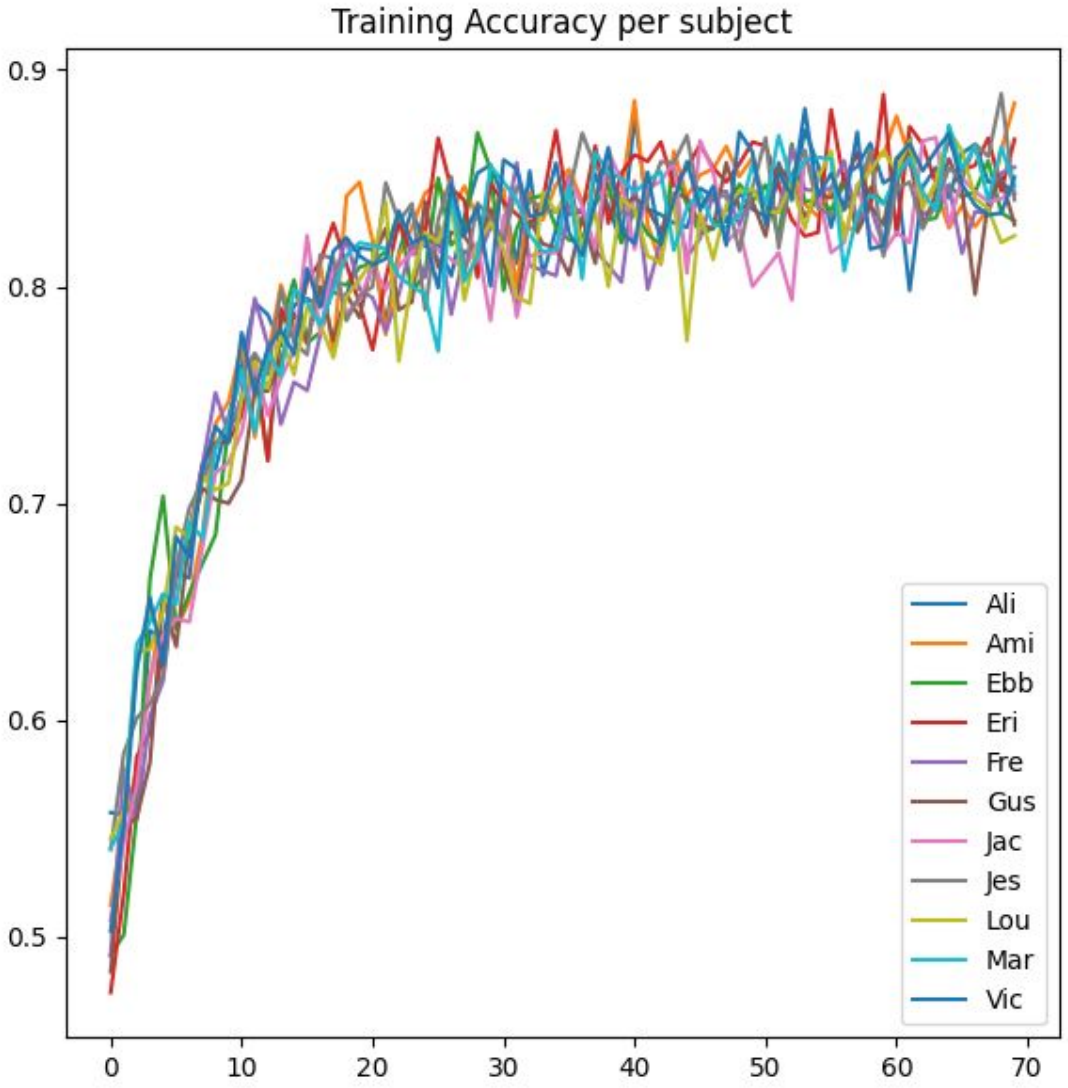
RegNet

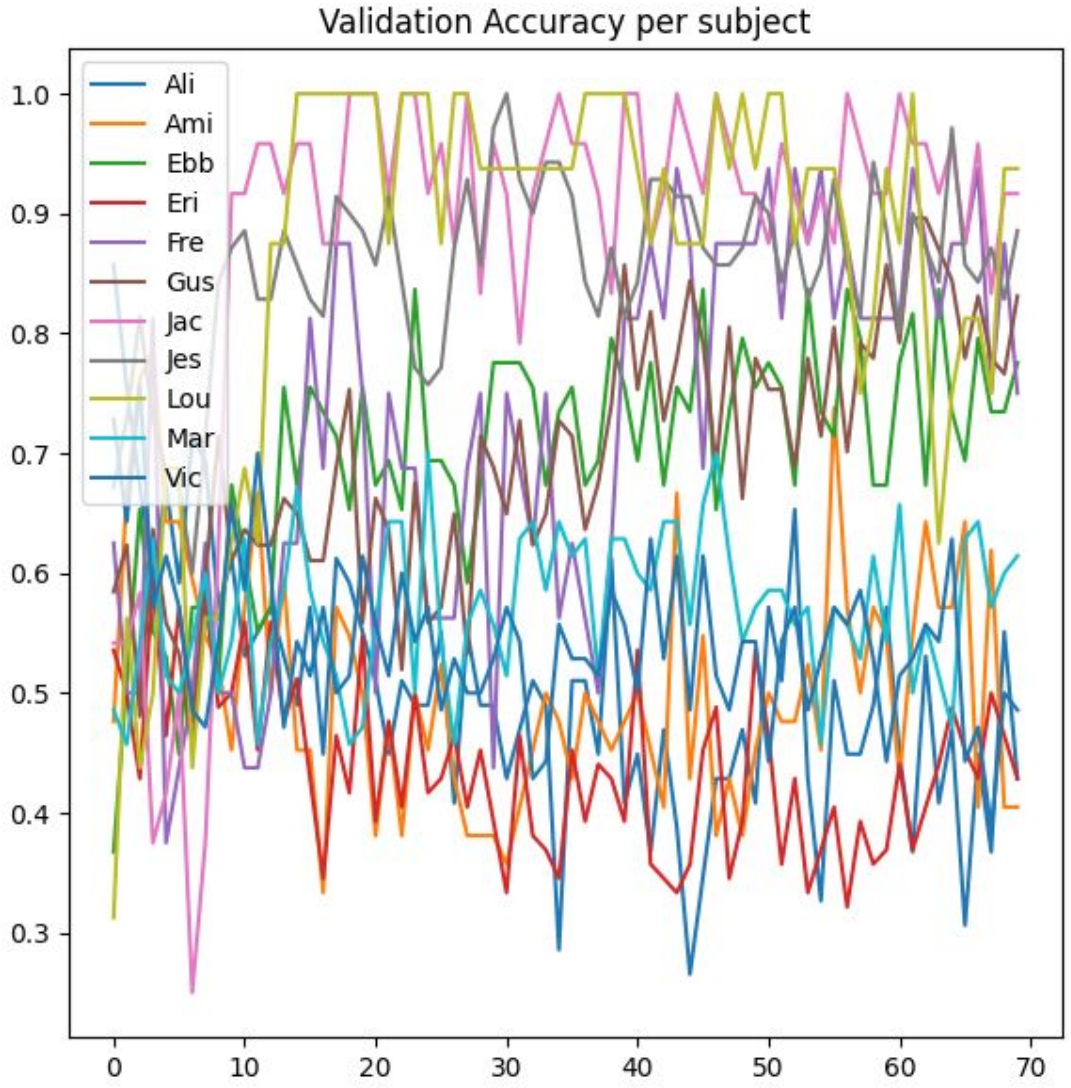
```
Ali  mcc: -0.3265986323710904 4
Ami  mcc: 0.0441510785688347 9
Ebb  mcc: 0.4
Eri  mcc: 0.0
Fre  mcc: 0.8181818181818182
Gus  mcc: 0.4803844614152613 7
Jac  mcc: 0.0
Jes  mcc: 0.1846372364689991
Lou  mcc: 0.0
Mar  mcc: 0.3340765523905304 7
Vic  mcc: -0.0330589802453643 14
```

CNN

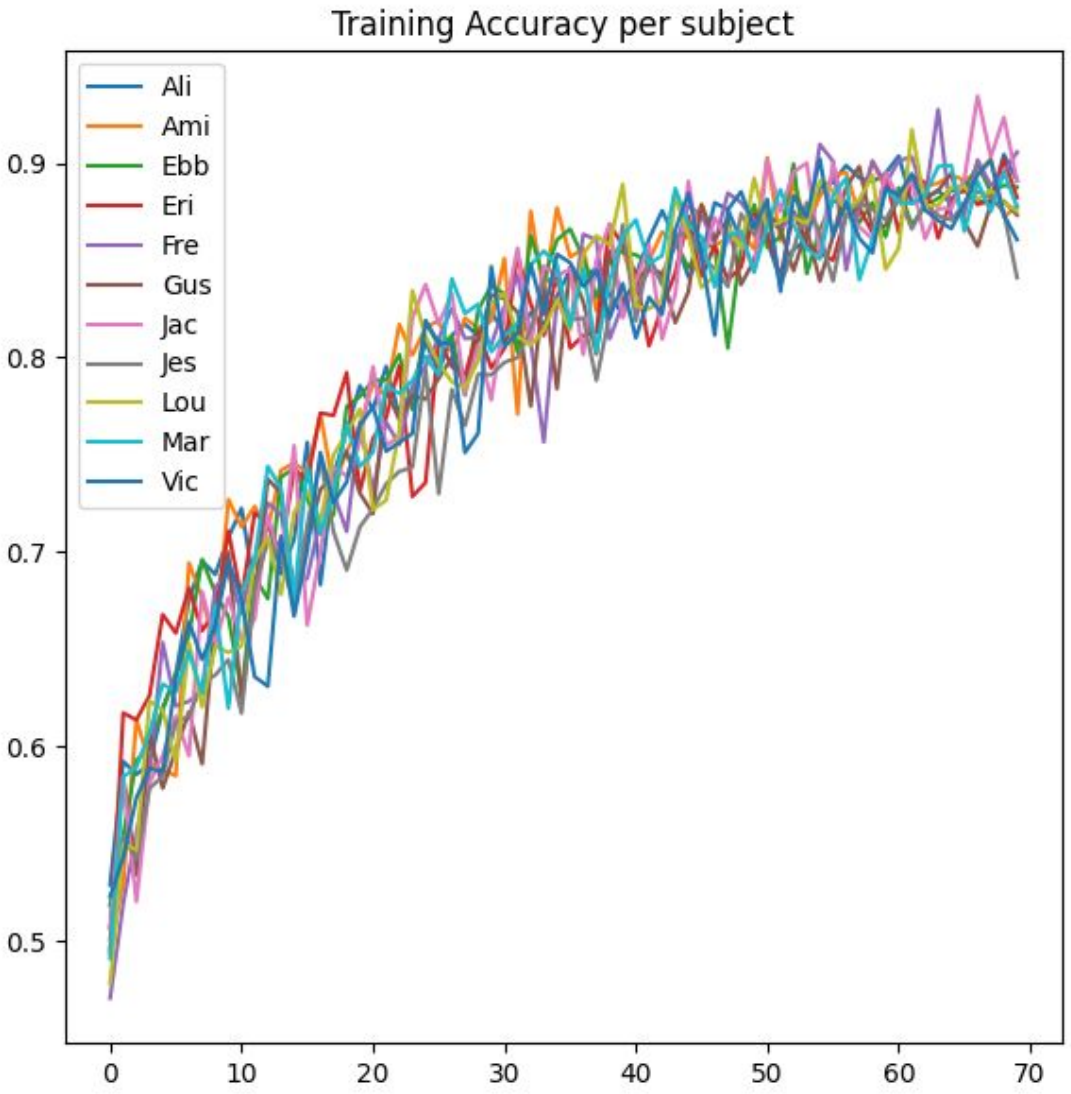
```
Ali  mcc: -1.0
Ami  mcc: 0.5270462766947299
Ebb  mcc: 0.3265986323710904 4
Eri  mcc: 0.2001978239850581
Fre  mcc: 1.0
Gus  mcc: 0.7745966692414834
Jac  mcc: -0.3015113445777636 3
Jes  mcc: 0.7924058156930615
Lou  mcc: 0.0
Mar  mcc: 0.1186781658193853 4
Vic  mcc: 0.3144854510165755 5
```

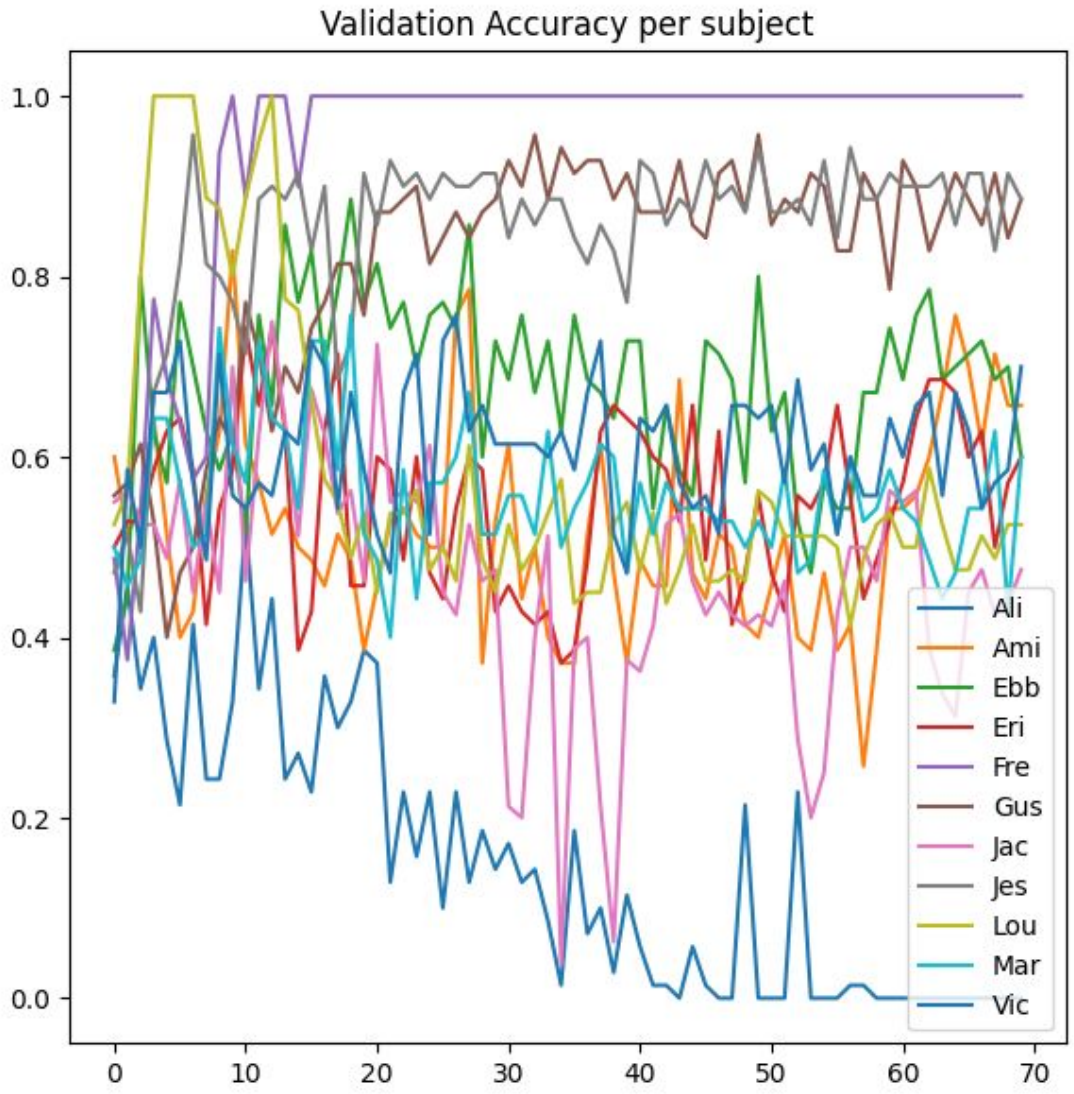
MaxVit


Training Accuracy per subject
0.9
0.8
0.7
0.6
0.5
0
10
20
30
40
50
60
70
Ali
Ami
Ebb
Eri
Fre
Gus
Jac
Jes
Lou
Mar
Vic


Validation Accuracy per subject
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
10
20
30
40
50
60
70
Ali
Ami
Ebb
Eri
Fre
Gus
Jac
Jes
Lou
Mar
Vic

CNN


Training Accuracy per subject
Ali
Ami
Ebb
Eri
Fre
Gus
Jac
Jes
Lou
Mar
Vic
0.9
0.8
0.7
0.6
0.5
0
10
20
30
40
50
60
70


Validation Accuracy per subject
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
70
Ali
Ami
Ebb
Eri
Fre
Gus
Jac
Jes
Lou
Mar
Vic

Cross-validation results

On select data

- CNN
  - Combined F1: **0.736**
  - Mean MCC: **0.604**
- RegNet
  - Combined F1: 0.665
  - Mean MCC: 0.429
- MaxVit (Hyper-parameter tuned using HyperOpt)
  - Combined F1: 0.667
  - Mean MCC: 0.443

MCC scores

MaxVit

```
Ami  mcc: 0.26413527189768715
Ebb  mcc: 0.5673665146135802
Eri  mcc: 0.353553390593273 73
Fre  mcc: 0.6123724356957945
Jes  mcc: 0.7924058156930615
Mar  mcc: 0.192450089729875 26
Vic  mcc: 0.319801074533415 66
```

RegNet

```
Ami  mcc: 0.5310850045437944
Ebb  mcc: 0.5927489783638191
Eri  mcc: 0.46068221272042836
Fre  mcc: 0.6123724356957945
Jes  mcc: 0.5196152422706631
Mar  mcc: 0.32090298129536804
Vic  mcc: -0.0294244943168249 85
```

CNN

```
Ami  mcc: 0.5270462766947299
Ebb  mcc: 0.6236095644623235
Eri  mcc: 0.39223227027636803
Fre  mcc: 1.0
Jes  mcc: 0.6324555320336759
Mar  mcc: 0.7888106377466154
Vic  mcc: 0.26650089544451305
```

Results over filtered dataset

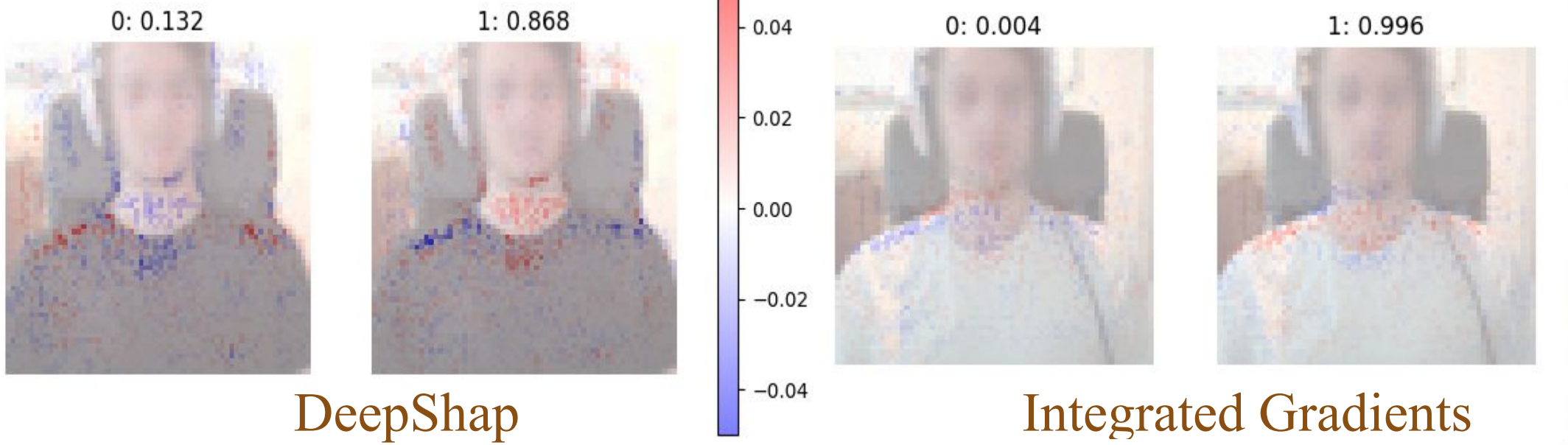
On Train-Test split data (Evaluated on the 3 most confusing subjects)

- CNN
  - Final F1: 0.82
  - Final MCC: 0.690
- RegNet
  - Final F1: 0.78
  - Final MCC: 0.577
- MaxVit (Hyper-parameter tuned using HyperOpt)
  - Final F1: **0.86**
  - Final MCC: **0.728**

DeepShap

Integrated Gradients

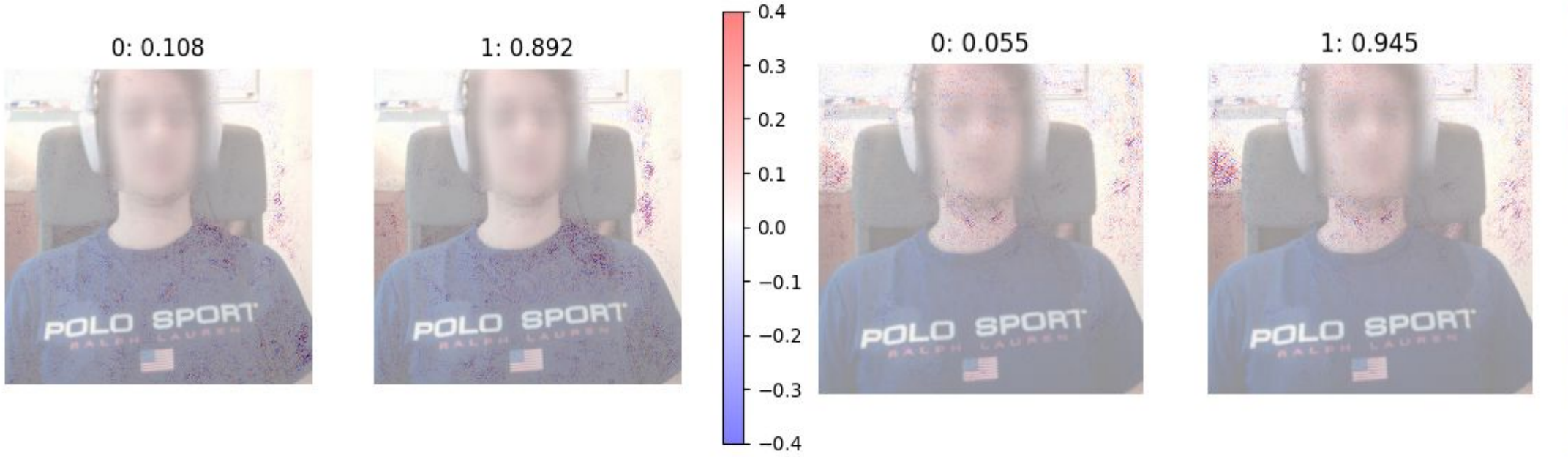
CNN



DeepShap

Integrated Gradients

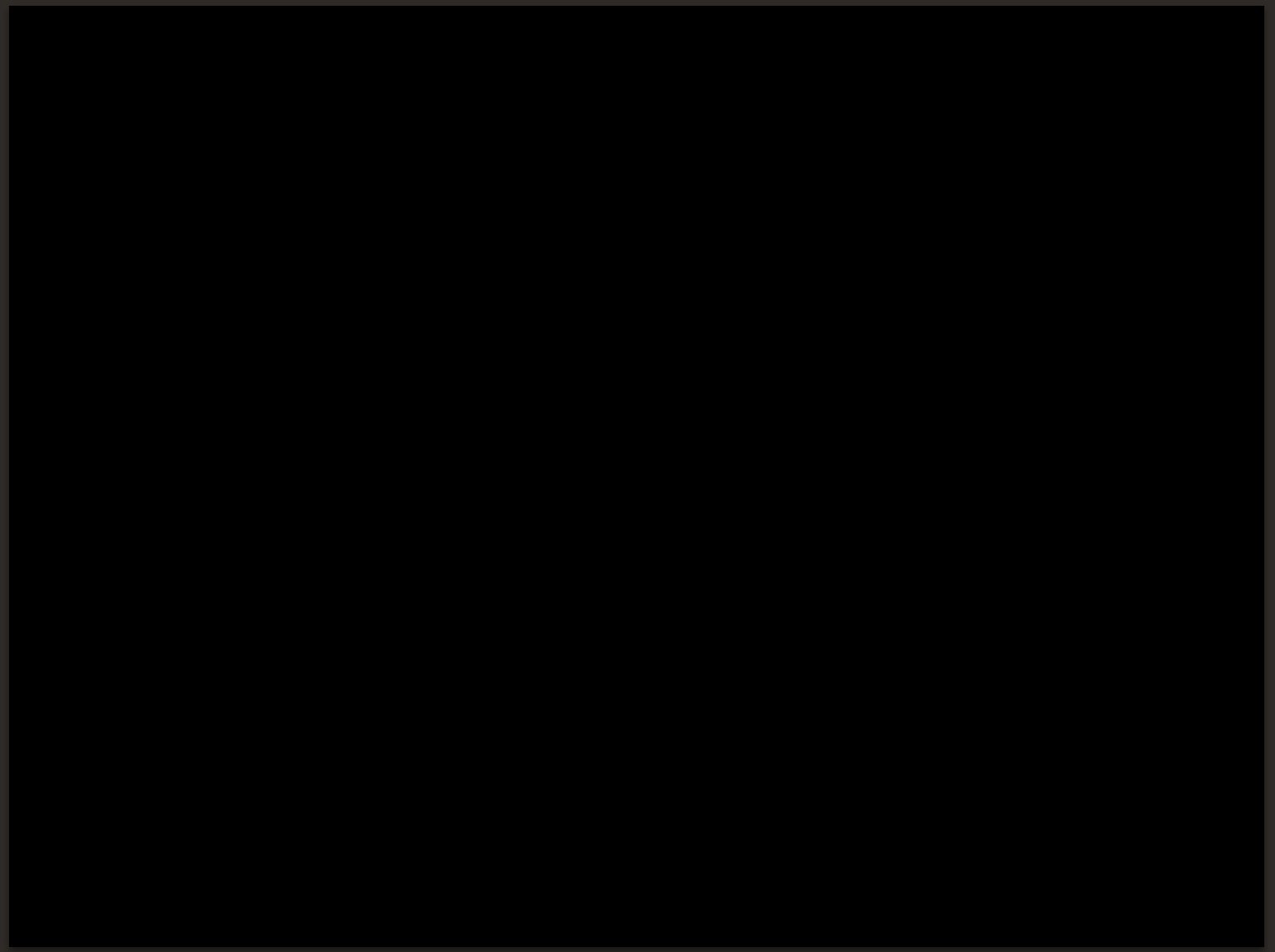
MaxVit



Conclusions

Conclusions

- Proof of concept indicated the task is learnable
- Adequate results with limited data and optimization
- Future work
    - Gather more diverse data
    - Fine-tune pre-trained model
    - Compile into usable background application



Questions

LUND
UNIVERSITY