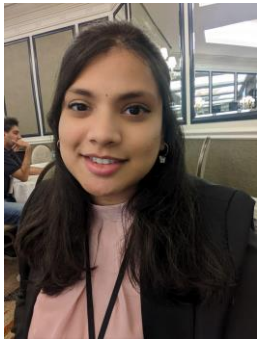
Vision-based Human Activity Recognition (HAR) using Transfer Learning Approach for Internet-of-Things (IoT) Applications

Team Introduction

Will Downey
M.Eng Student
Computer Science

Srijeet Halder
Ph.D. Candidate
School of Construction

Ashit Harode
Ph.D. Candidate
School of Construction



Nikitha Chandrashekar
Ph.D. Student
Computer Science

Problem Description

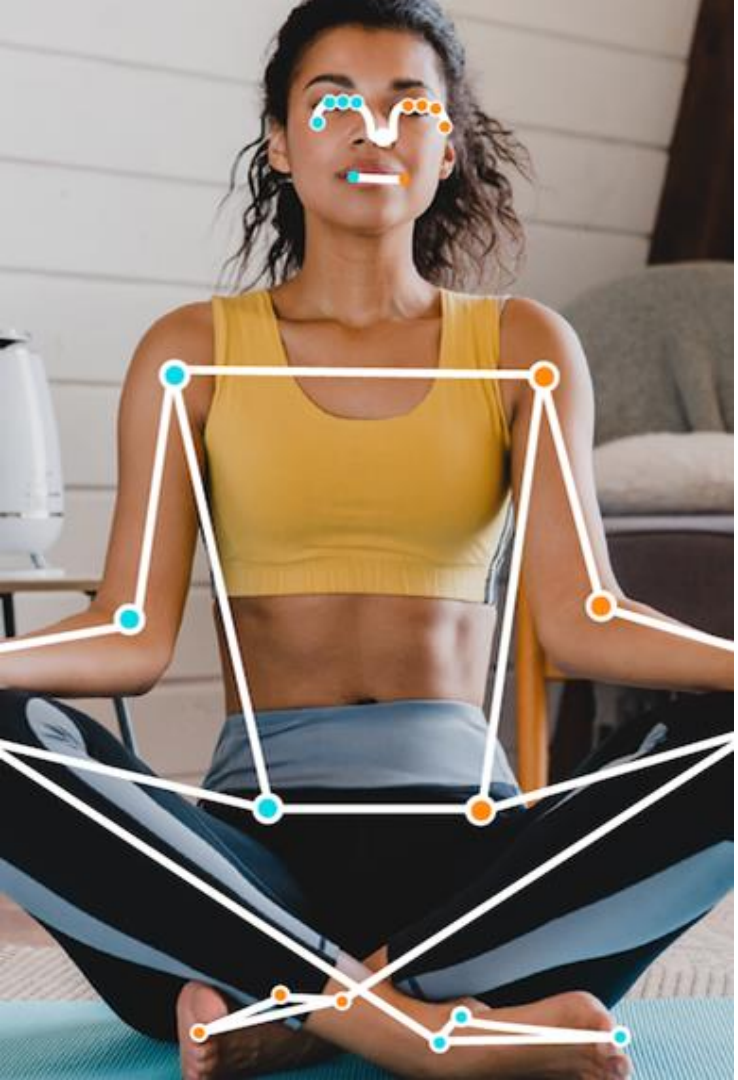
- Smart Homes and Buildings (SHaB) improve the user's quality of life and resource usage.
- Kind of activity influence thermal behaviour of the space.
- Strong autonomous HVAC system considers this as a factor.
- Use a vision-based machine learning model to classify images as active or passive

Dataset

- 12,000+ labeled images featuring 15 different classes of human activities.

Pre-Processing

- Mapped 15 classes into 2 classes
- Convert the multi-class classification problem to binary classification.



Mediapipe Pose

- ML solution for high-fidelity body pose tracking.
- Infers 33 3D landmarks from RGB video frames.
- Locates person/pose ROI in the frame.
- Predicts pose landmarks within ROI.

Data Augmentation

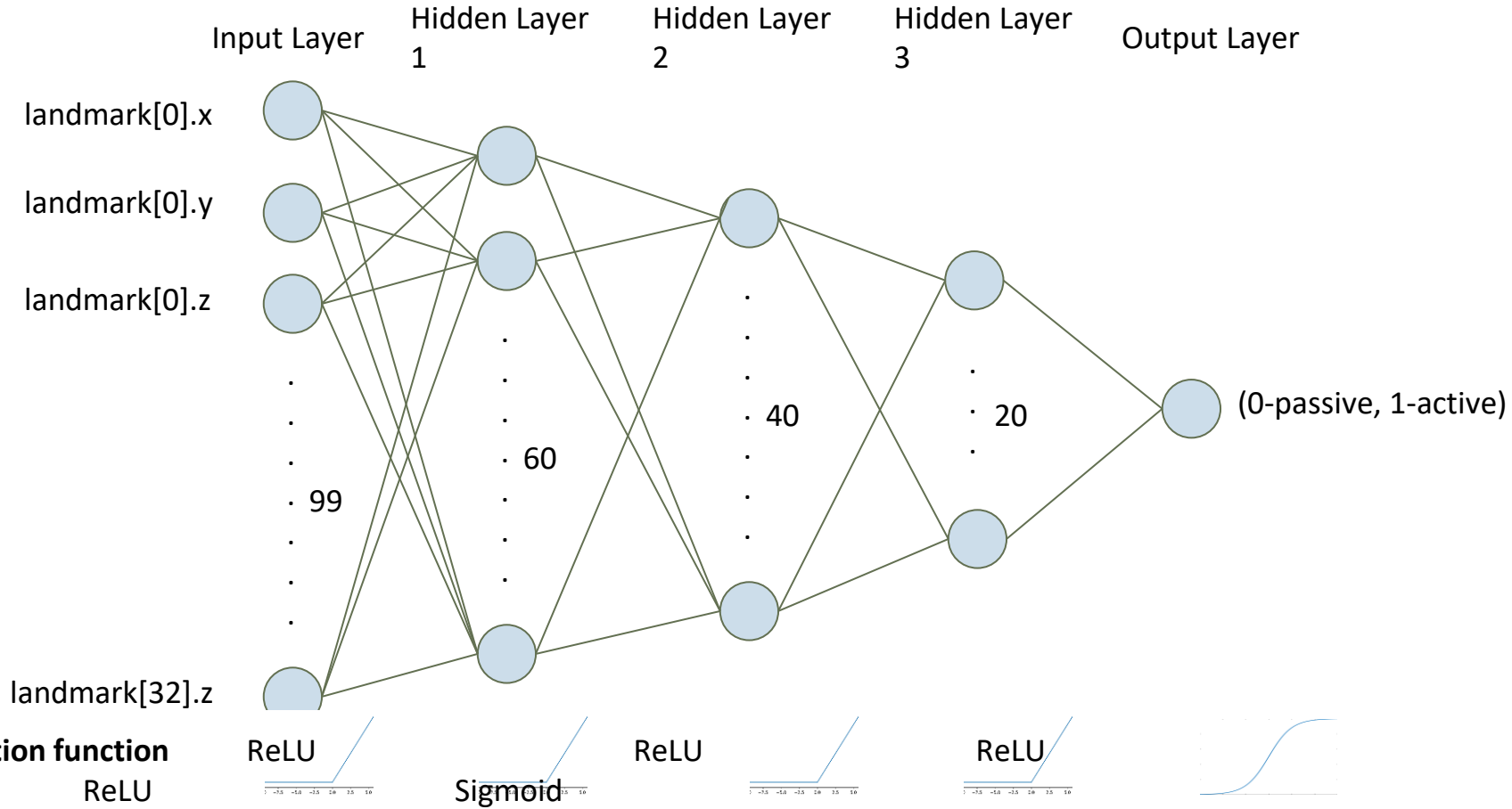
+random_rotation

+random_resize

+random_flip

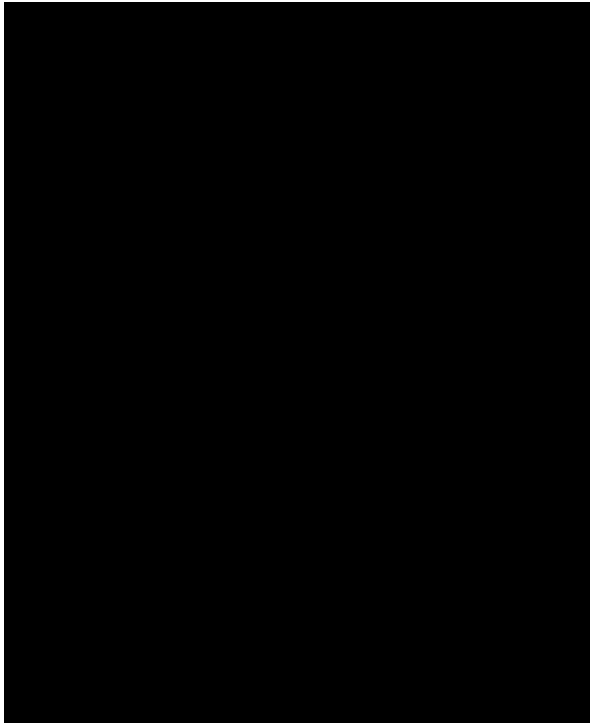
augment_image(image)

Model Training



Model Training

- Training framework = Tensorflow + Keras
- Loss function = Binary Cross-Entropy
- Optimizer = Adam
- Batch size = 32
- Epochs = 22 (early stopping)
- Train/Test split = 80/20



Training Results

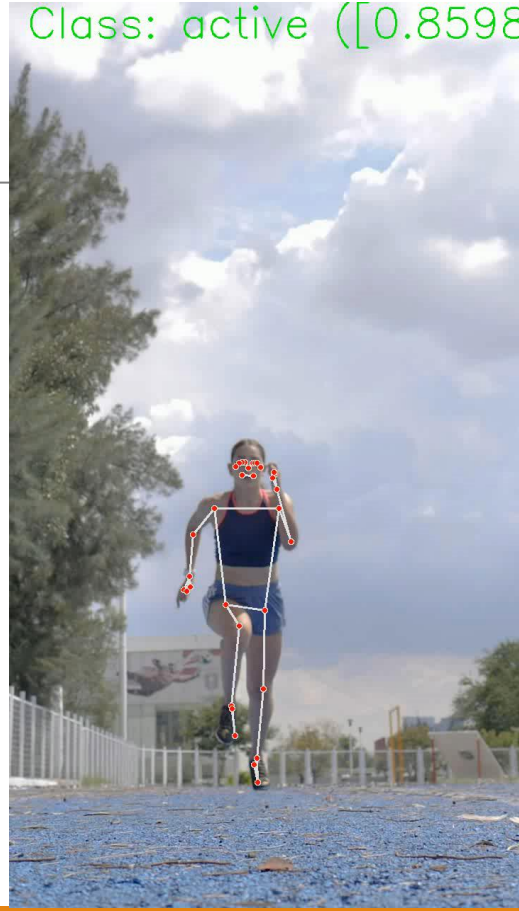
- Results after running for 22 epochs:
  - Precision: 0.79
  - Recall: 0.79
  - F1-Score: 0.79
  - Accuracy: 0.79

```
51/51 [==============================] - 0s 2ms/step
              precision    recall  f1-score   support

           0       0.74      0.75      0.74       658
           1       0.83      0.82      0.82       972

    accuracy                           0.79      1630
   macro avg       0.78      0.78      0.78      1630
weighted avg       0.79      0.79      0.79      1630

accuracy: 0.7907975460122699
```

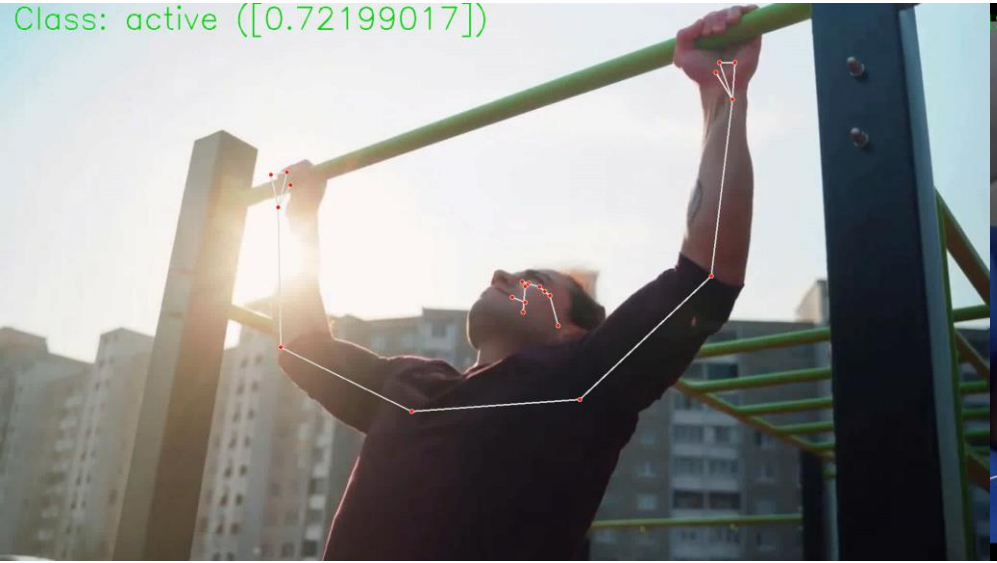
Evaluation on External Video

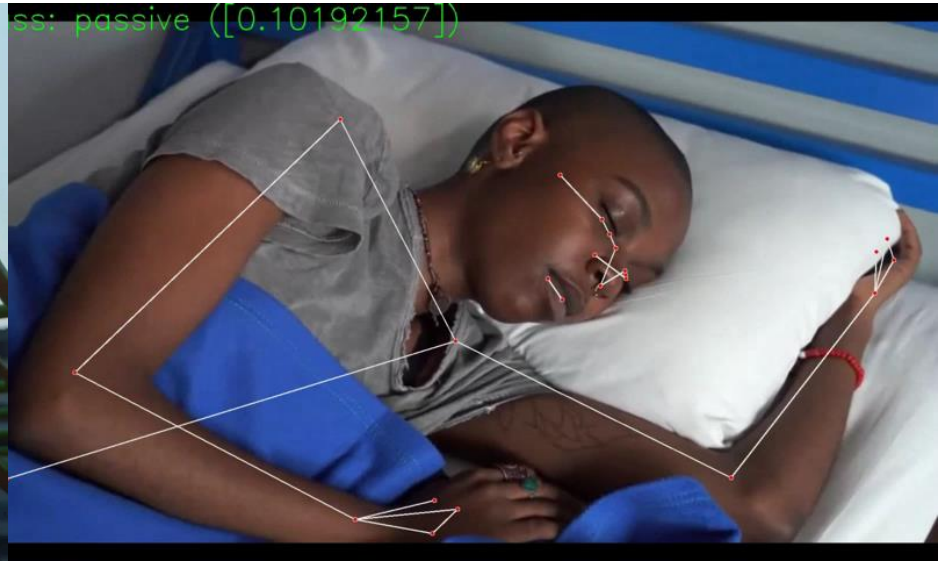
- Testing on shown video outside of data set
- Results:

```
8/8 [==============================] - 0s 9ms/step
              precision    recall  f1-score   support

           0       1.00      0.97      0.98       242
           1       0.00      0.00      0.00         1

    accuracy                           0.96       243
   macro avg       0.50      0.48      0.49       243
weighted avg       0.99      0.96      0.98       243

accuracy: 0.9629629629629629
```



Evaluation on External Video





Lessons Learned

- Neural Networks are powerful ML techniques that can extract hidden patterns
- More data = improved accuracy = more training time
- AI models require extremely large amounts of data to be effective
- Data augmentation can be a useful technique to increase size of training dataset

Future work

- An input-action mapping needs to be created that performs certain action on the HVAC system based on the recognized human activity