# **Action Recognition**

Problem Statement

Given a set of images of a video the aim of the work is to

Analyze the different actions in the video using Deep learning architecture.

The performance of the work is evaluated using performance

Parameters namely precision

Dataset Details

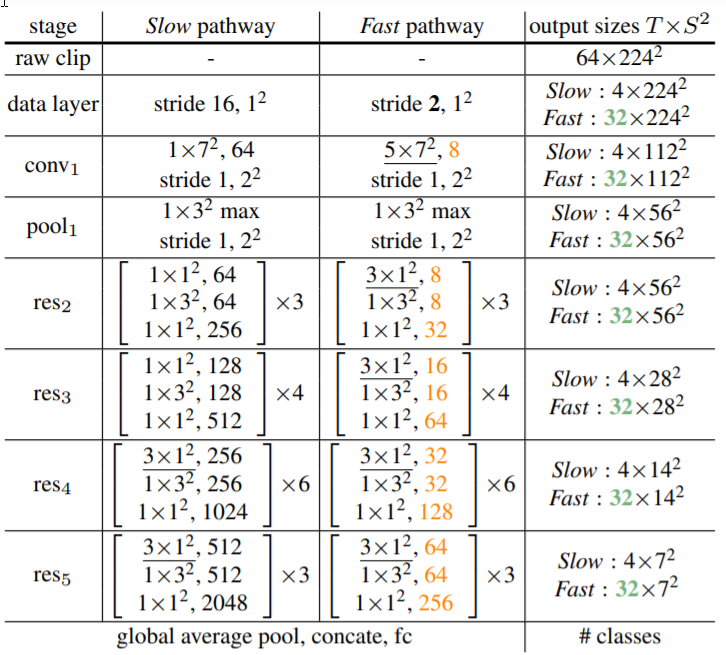
The dataset contains 400 human action classes, with at least 400 video clips for each action. Each clip lasts around 10s and is taken from a different YouTube video. The actions are human focussed and cover a broad range of classes including human-object interactions such as playing instruments, as well as human-human interactions such as shaking hands. We describe the statistics of the dataset, how it was collected, and give some baseline performance figures for neural network architectures trained and tested for human action classification on this dataset. We also carry out a preliminary analysis of whether imbalance in the dataset leads to bias in the classifiers.

Description of the Scene

Provide details of the scene with some sample images and indicate the type of actions.

|  |  |  |  |
| --- | --- | --- | --- |
| Action Name | Description | Sample Image | Output |
| Climbing a tree | A person is climbing a tree by stepping on its branches |  | [climbing tree] 0.994.  [trimming trees] 0.004. |
| Riding a bicycle |  |  | [motorcycling] 0.809.  [riding a bike] 0.145.  [riding unicycle] 0.012. |
| Playing a piano |  |  | [playing recorder] 0.658.  [shuffling cards] 0.046.  [playing trumpet] 0.030. |
| Javelin Throw |  |  | [javelin throw] 0.998.  [pole-vault] 0.001. |
| Riding a scooter |  |  | [riding scooter] 0.895.  [hoverboarding] 0.105. |

Deep Learning Architecture Details



SlowFast networks for video recognition. This model involves

1. A Slow pathway, operating at low frame rate, to capture spatial semantics, and

Aa Fast pathway, operating at high frame rate, to capture motion at fine temporal resolution. The Fast pathway can be made very lightweight by reducing its channel capacity, yet can learn useful temporal information for video recognition.

Graphs and Discussions

From the results we can infer that the actions are being recognized on the abstract level.

SlowFast model gives an mAP of 42.5 for the kinetic400 dataset. The accuracy for each sample images has been shown in the table above.

References [Web References URL]

1. SlowFast model - <https://arxiv.org/pdf/1812.03982.pdf>
2. Kinetic400 - <https://arxiv.org/pdf/1705.06950.pdf>