Model used:

The model we’re using for human activity recognition comes from Hara et al.’s 2018 CVPR paper, [***Can Spatiotemporal 3D CNNs Retrace the History of 2D CNNs and ImageNet?***](https://arxiv.org/abs/1711.09577)

In this work the authors explore how existing state-of-the-art 2D architectures (such as ResNet, ResNeXt, DenseNet, etc.) can be extended to video classification via 3D kernels.

The authors argue:

* These architectures have been successfully applied to image classification.
* The large-scale ImageNet dataset allowed such models to be trained to such high accuracy.
* The Kinetics dataset is also sufficiently large.

**…and therefore, these architectures should be able to perform video classification** by (1) changing the input volume shape to include spatiotemporal information and (2) utilizing 3D kernels inside of the architecture.

Data set used:

The dataset our human activity recognition model was trained on is the [**Kinetics 400 Dataset**](https://arxiv.org/abs/1705.06950).

This dataset consists of:

* 400 human activity recognition classes
* At least 400 video clips per class (downloaded via YouTube)
* A total of 300,000 video

To learn more about the dataset, including how it was curated, be sure to refer to Kay et al.’s 2017 paper, [***The Kinetics Human Action Video Dataset***](https://arxiv.org/abs/1705.06950).

Reference:

<https://pyimagesearch.com/2019/11/25/human-activity-recognition-with-opencv-and-deep-learning/>