Introduction:

Machine Learning

Machine learning is a sub-field of artificial intelligence study that is focused on creating programs for a computer to autonomously make decisions about new data based on a model from data it has already been given.

Many common machine learning models try to emulate the natural way neurons work in organic brains, and so are called neural networks. In artificial neural networks, the inputs are given different weights, and a small calculation is run with them to determine the output.

Convolutional Neural Networks (CNN)

CNNs are a form of machine learning algorithm commonly used for analyzing visual data. Like other neural networks, they use multiple hidden layers to perform a variety of transformations on the data to categorize it. They take advantage of the fact that the input will be an image, and arrange the neurons in three dimensions (width, height, depth).

The convolutional layers of these networks take a small filter and move it across the input. At each position, the dot product is calculated between the filter and the region it covers, then this value is put in another array, which is passed to the next layer.

We took some inspiration for our network architecture from some well-known networks such as AlexNet and ImageNet. In our research we worked to classify videos from the UCF101 Action Recognition Data Set.

Video Classification

Video classification is different from image classification, because in addition to the spatial data (a single frame), the temporal data (difference between frames) must also be taken into account. This can sometimes be a little complicated, as some video formats run at different frame rates; however most are 20-30 frames per second. Also, it is important to be careful of things such as a change in camera angle, or a transition between shots or scenes.