The goal of our research was to design a CNN for action recognition in video clips. We wanted to know if selecting a certain number of frames from each video clip and adding their pixel values together at the start of a CNN would both improve accuracy and minimize loss more effectively than contemporary models.

Machine Learning

Machine learning is a field of artificial intelligence study that is focused on programming computers to make decisions about new data based on a model from data it has already seen. Many machine learning models try to emulate the way neurons work in organic brains, and so are called neural networks.

Convolutional Neural Networks (CNN)

CNNs are a form of neural network commonly used for analyzing visual data. They use 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). [1]

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. [2]

References:

[1] A. Karpathy, "Convolutional Neural Networks: Architectures, Convolution / Pooling Layers," *github.io*, para. 2, pub. date unavailable. [Online]. Available: http://cs231n.github.io/convolutional-networks/. [Accessed June 13, 2019].

[2] A. Karpathy, G. Toderoci, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei, "Large-scale Video Classification with Convolutional Neural Networks,"

i don't have what journal [2] was pulled from

[3] A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *Communications of the ACM*, Vol. 60 No. 6, Pages 84-90, June 2017.