A basic 2D convolution slides a small window over a 2D image, and calculates the dot product between the window and the corresponding input patch at every position. For higher dimensional input, depthwise convolution slides a separate filter along the image’s dimension for each channel. Depthwise separable convolution does the same thing and applies an 1x1 convolution across all channels. These two methods process higher-dimensional input while keeping the number of operations low. I3D and I2D models repeatedly use convolution and pooling layers to condense video frames into smaller vectors. CNN can also be used in other areas like speech and NLP using techniques.