**MODEL ARCHITECTURE**

* The frames of each video are converted to NumPy arrays
* These arrays are then given as input to the standard VGG16 pre-trained model (excluding the last layer) to produce an encoded form of train data. It is expected that the features of the images are learned and encoded during this process.
* The results of the above are used as inputs to our custom model with the following architecture:

