**Preprocessing**

The videos are fed into a preprocessing pipeline that is outline in the following major steps:

1. Extracting the subject’s face from the Images in the video. (and verify that the extracted image is a face)
2. Sampling the extracted faces sequence into a fixed size through down sampling (frame skipping) to get sequences of similar length to feed to the neural networks. (sampling is done after face cropping because not all frames successfully yield a face image after cropping)
3. Resizing the final sequences to a fixed width and height to ensure the input structure is constant over all samples.

**Model Training**

After preprocessing, the sampled and normalized images are fed into a 3D CNN network to extract features from the videos, finally these features are fed to fully connected layers to further process the data and discover more relations, and at the end there is a fully connected layer followed by one neuron in the output layer that outputs the probability of deception in the input video.

In this approach, feature extraction is done in the 3DD CNN layers.

**Important remarks and discoveries**

On using a random data split (this causes unreal results because certain subjects appear so many times in the dataset so the models can potentially learn to identify these subjects rather than detect lies) The accuracy yielded is very high reaching 95% specially using a CNN-LSTM model, although it’s a bit less with a 3D CNN at around 85%, but surprisingly when manually splitting the data to ensure that the test set contains videos of subjects that have never appeared in the train set (to eliminate all bias superstitions) the CNN-LSTM achieves poor results with accuracy below 60% while the 3D CNN maintains decent results with peek accuracy of 83% on both train and test sets.