You should first run the ouluPreprocessingScript.m in Matlab in order to generate the dataset in the required format. The dataset should be saved to "oulu\data" and contains the following variables:

dataMatrix: matrix which contains all the frames flattened, size = noFrames x length of image vector (=image height X image width)

targetsVec: contains the targets per frame, size = noFrames x 1

filenamesVec: cell array which contains the filename that corresponds to each sequence, size = noSeq x 1

iterVec: contains the iteration (each utterance is repeated 3 times) that corresponds to each sequence, size = noSeq x 1

subjectsVec: contains the subject ID that corresponds to each sequence, size = noSeq x 1

targetsPerVideoVec: contains the target that corresponds to each sequence, size = noSeq x 1

videoLengthVec: contains the length of each sequence, size = noSeq x 1

dataMatrixCells: cell array which contains the frames of each sequence grouped in cells, size = noSeq x 1

For the python code to run, only the following variables are needed: **dataMatrix, targetsVec, subjectsVec, videoLengthVec**.

Then you should run the preTrainEncoderWithRBMs.m in Matlab (you will need the following toolbox *https://github.com/stavros99/DeepLearningToolbox_Matlab*) in order to pre-train the encoder for each view. The weights and biases should be saved to "oulu\models". This step is necessary if a new set of pre-trained weights and biases is needed. Pre-trained weights and biases for each view are already available at *https://ibug.doc.ic.ac.uk/resources/EndToEndLipreading/*.