Dataset is collected and unzipped and pre-processing of image takes place