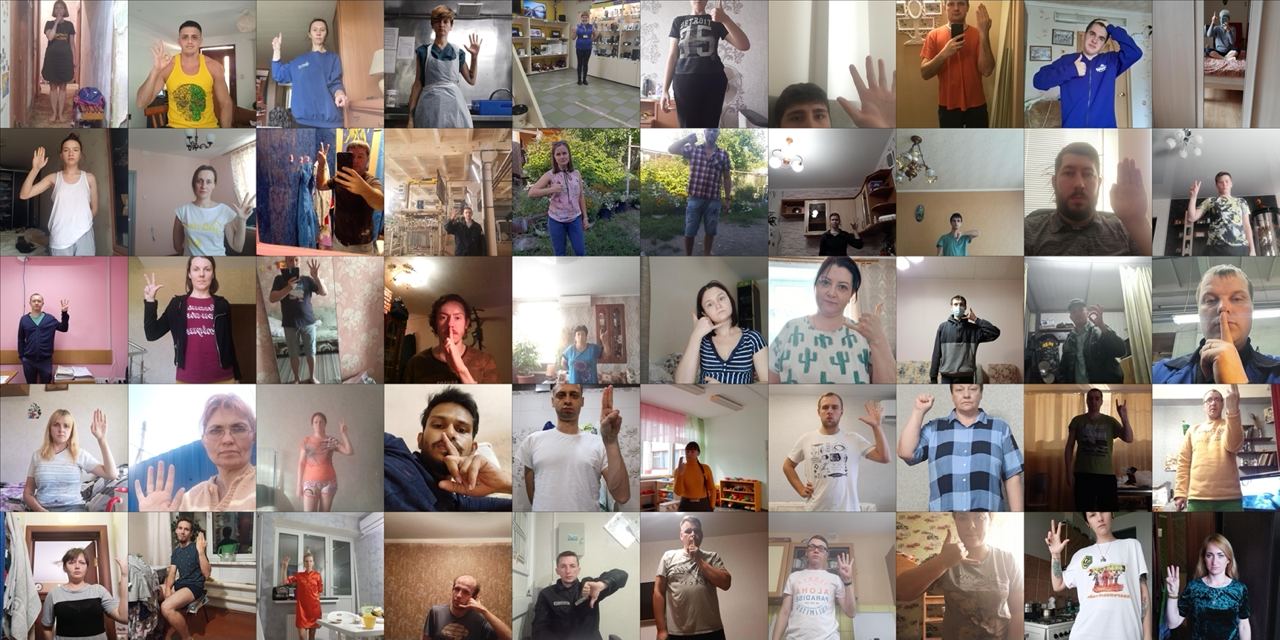
# HaGRID

The dataset used for this thesis is HaGRID (Hand Gesture Recognition Image Dataset). It contains 552’992 Full HD (1920 x 1080) images which amounts to 716 GB of data. The images are originally divided into 18 classes of different gestures, however, for the sake of this project they are all merged into a single “hand” class. The training-test split is by default 92%/8%. The dataset contains over 34’000 different people from various age groups and ethnicities. The scenes range from dark indoor environments to bright outdoor settings. The subjects were instructed to show different gestures, mostly in third person, with a 0.5m to 4m distance to the camera. Figure X shows a subsample of images from HaGRID. 

The dataset HaGRID seems to be an optimal choice for the task of hand recognition in the context of sign language detection for multiple reasons. First of all, the dataset is enormous and comes with a great variety of different scenes. This is substantial for the training of a reliable and well performing model. Second, most of the images are captured in third person. Consequently, the dataset is a good fit for the use case of sign language detection where target subjects are also going to be presented in third person to the camera. Lastly, the variety in gestures will make the trained network more robust to variables such as finger occlusion and hand position as this is also the case for sign language.

Figure X shows a sample of an image from HaGRID together with its JSON annotation. The annotation compiles bounding box coordinates and gesture classification of each hand. Furthermore, each person is denoted with a unique user ID and there is a distinction between left and right hand.

