**1 slide)** It is well known that a person is able to divine the behaviour of people without effort or delay. But for artificial intelligence it is not so simple task. So…

**2 slide)** The purpose of my work is to develop the robust algorithm of facial expression recognition and smile detection on face image based on a relatively new approach in the field of deep machine learning - convolutional neural networks.

**3 slide)** An automatic solution of this problem can be used in clinical psychology and psychiatry in lie detectors. In recent years, the use of such algorithms is becoming popular in the field of video analytics, such as evaluating front office staff performance or in security systems. Also similar solutions can be used for the retail in the field of entertainment services.

**4 slide)** The network’s architecture was implemented using Caffe framework. This library allows to describe the convolutional neural network by means of ready algorithms of machine learning.

To accelerate the neural network operation, the training and testing processes were performed parallel, on a large number of independent streams on GPU. For this, the technology NVIDIA CUDA was used. This technology is cross-platform and is supported by all modern NVIDIA graphics cards.

Also to solve some auxiliary tasks there were written some scripts on Python.

**5 slide)** The developed neural network consists of 4 convolutional layers, 4 layers with the ReLU activation function, 4 layers realizing the local normalization process, 3 layers realizing the maxpooling operation, one fully connected layer and one softmax layer. As a numerical optimization algorithm, a stochastic gradient descent was chosen.

Patches of the size 122×122×5 pixels are randomly cropped from grayscale images of resolution 128×128 pixels. They are the input of this neural network. On the first   
convolutional layer, 10 kernels of the size 3×3×5 are applied to the patches with a pixel step of 1. After applying the ReLU activation function, local normalization and maxpooling operation with a 2×2 filter and pixel step of 2, the features will have 10×60×60 size. A similar sequence of operations is performed one by one for 20 kernels of 5×5×10 size, 40 kernels of 7×7×20 size and 80 kernels of 11×11×40 size. After all the operations done, a vector of 80 is transferred to the fully connected layer. This layer allows to classify the input image to the certain facial expression. The learning and testing processes for the deep convolution neural network lasted around 7 minutes.

The pictures on every maxpooling layer are shown on the bottom of this slide.

**6 slide)** Numerical experiments for developed algorithm were performed for the images of Multi-Pie database. This database contains more than 750,000 color images from 337 subjects. The images were made on different angles (less than 90°) with different lighting of the scene.

The scheme of camera location is shown on the right of this slide.

**7 slide)** In Multi-Pie database there are images of 6 different types of emotions, such as neutral, smile, surprised, squint, disguist and scream.

**8 slide)** The sampling for numerical experiments contains 30000 randomly selected images with different lighting of scene and angle of view less than 45°. These images were labeled according to 2 or 6 classes depending on the problem . Also on each picture from the sampling there was cut the face image of 128×128 size and transformed into black-and-white mode. Such conversion was implemented by means of PICO algorithm.

**9 slide)** The main result of numerical experiments is shown on this slide. For every problem accuracy is more than 85%.

**10 slide)** On this slide, for each problem you can see the dependency of accuracy and number of iterations, when the classifier is stopping. After 10000 iterations the values of accuracy doesn’t increase.

**11 slide)** On this slide there are presented dependencies of values of loss function and numbers of iterations. For each problem you can see, that the developed algorithm converges.

**12 slide)** For smile detection problem there was given a confusion-matrix, which allow to evaluate the quality of classifiers, and values of metrics, such as precision, recall and F-score.

As you can see, type 1 and 2 errors are insignificant. The values of F-score for each of these classes are 0,95.

**13 slide)** Also for smile detection problem there was built ROC curve and calculated the value of AUC-ROC. It is approximately equal to 0,98.

**14 slide)** On this slide there is presented a confusion-matrix for the problem of facial expression recognition. The worst classified facial expression are “Disguist” and “Squint”. The classifier often confused them.

**15 slide)** This fact is also confirmed by metrics. The values of precision, recall and F-score for these classes are less than others.

**16 slide)** It can be explained as follows: in Multi-Pie the images with these emotions were hardly distinguishable. However, this developed convolution neural network copes with facial expression recognition by images well.

**17 slide)** But, of course, the developed convolutional neural network can be and will be improved due to more qualitative samples, using other databases, for example CK, CK+. To increase the robustness of realized algorithm there will use linear transformation for images. These transformations are the horizontal rotation, skewing the center and zooming with cropping at four corners of the image. At this moment, numerical experiments are implementing. Thereafter it is planned, that this algorithm will be plugged in cameras and tested on real images.