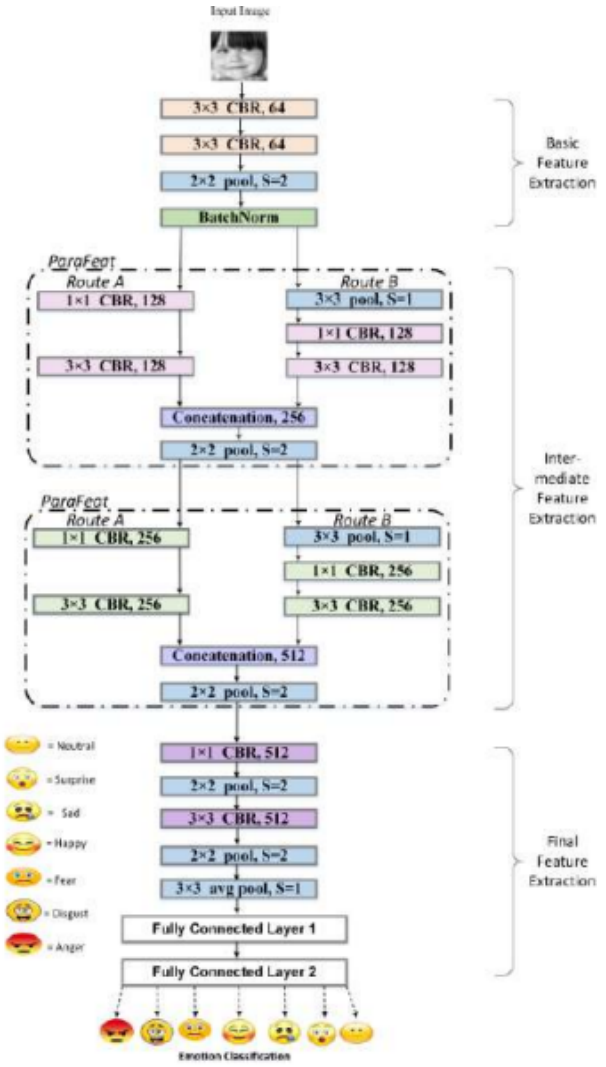

Figure 10. The structure of eXnet (original image from [19])

640x480 pixels. It should be noted, however, that the classes for this dataset differ from the above-mentioned ones (for example, instead of a neutral emotion, there is an emotion of despisal). This was one of the reasons why there was a need to form a combined dataset;

2) The "Students and colleagues" dataset. This dataset was collected from data obtained by self-determined image collection and processing. It is a set of images of the faces of 38 people with the expression of basic emotions. The training part included 20,947 images, the control part – 2,101;

3) Internet Emotions. This dataset was formed from images taken from the Internet and includes one-man photos of 295 people, which include training (268 images) and control (27 images) datasets.

Thus, the total size of the training dataset was 25,830 images and the control one – 2,682.

As a result of additional training of the eXnet model on the dataset described above, the results presented in table I were obtained.

avg valid – the percentage of successfully recognized emotions taken as 1 in the test dataset.

Table I
RESULTS OF TRAINING THE EXNET MODEL

	avg. valid	*avg. valid softmax*
CK+	0.877	0.859
Students and Colleagues	0.765	0.742
Internet	0.37	0.305

avg softmax – the percentage of successfully recognized emotions, taking into account the obtained probability from the range (0..1] in the test dataset.

ck+, **student_colleague** and internet denote the overall accuracy that results from the corresponding test datasets. **average_valid** and **average_valid_softmax** are the metrics used to evaluate accuracy.

F. Semantic analysis

The described integration mechanism allows enriching the knowledge base with the results of recognition of various models used by the computer vision module (identification model and emotion recognition model). The processing of this knowledge will be no different from the processing of any other knowledge in the ostissystem, regardless of whether they got there from the computer vision module, any sensors, visual or natural language interface or in some other way. In this case, computer vision is another receptor of the system

Knowledge processing in the KB, i.e., semantic analysis, is performed by the problem solver. The problem solver is a set of agents that react to events in the knowledge base (for example, a problem definition), solve its problem (generating, transforming knowledge, accessing external systems) and put the result of the work in the same KB.

For example, one of the methods of knowledge processing can be the usage of logical inference [21], which generates new knowledge based on a set of rules. Logical rules, in the simplest case, can be represented by "if-then" bindings, where the "if" part describes the knowledge that must be in the knowledge base to make it possible for us to generate the knowledge described in the "then" part. The origin of such rules can be different: from adding them manually by knowledge base engineers to automatically generating them.

In the considered implementation of the hybrid system, the logical rules [21] are used to generate some standard system responses to the interlocutor's messages. These rules use such knowledge as the identification of the interlocutor and their current emotion. Figure 11 shows a fragment of such a rule in a simplified form for clarity (in a real system, such rules have a more complex specification).

The meaning of the rule is as follows: if we received a greeting message from a user, whose emotion is recognized by the system as "sadness" and whose name the