

BABEȘ-BOLYAI UNIVERSITY CLUJ-NAPOCA
FACULTY OF MATHEMATICS AND COMPUTER SCIENCE
SPECIALIZATION COMPUTER SCIENCE

DIPLOMA THESIS

**UltraLight-AttentionNET:
A LIGHTWEIGHT NEURAL NETWORK FOR
REAL-TIME AUTOMATED FACIAL EMOTION
RECOGNITION**

Supervisor
Prof. Dioșan Laura

Author
Tofan Cristian-Raul

2024

ABSTRACT

Automatic facial emotion recognition holds immense commercial potential but comes with a vast amount of challenges. It proved to be very useful in domains like: e-learning, entertainment, medical, criminal interrogation and driving systems. Observing this huge potential, a lot of researchers made efforts to solve this challenging task using different approaches, which can be divided into two main groups: classical approaches and neural network based approaches. Many of these solutions are very costly in terms of computational cost and storage, which makes them impossible to be used in a real-time prediction system or on embedded devices.

This paper aims to solve the problem of automated facial emotion recognition in real-time, but also with a reduced storage need. Multiple experiments were conducted with various architectures in order to achieve great results in terms of computational cost, generalization ability, and the size of the file that stores the model. Three of these architectures are presented in this paper, all of them being able to classify an image in less than 0.1 seconds. In order to achieve that, the architectures combine efficient attention modules with depth-wise separable convolutions, standard convolutions and global average pooling instead of fully connected. Also, the proposed models achieved great results when they were compared to the state-of-the-art neural networks existing solutions.

The best architecture of these three is named UltraLight-AttentionNet and can be stored in a file of size 1.44 MB. It also proved that it can classify an image in an average of 0.08 seconds. Not only that it is efficient as computational cost and storage, but it also proved to have a good generalization ability. It was trained on the FER-2013 dataset, which is one of the most challenging datasets known in the facial emotion recognition domain and achieved 68.43% accuracy, which is better than the human accuracy achieved on the same dataset and also better than other light-weight neural network solutions.

Due to the great results achieved by the UltraLight-AttentionNet, it was integrated into an intelligent system that can predict emotions in a live video using a webcam. The intelligent system, named REmo, offers users the chance to predict their emotions in real-time on their computers.