An Embedded Continual Learning System for Facial Emotion Recognition

Olivier Antoni[1], Marion Mainsant[1], Christelle Godin[2], Martial Mermillod[3], and Marina Reyboz[1]

[1] Univ. Grenoble Alpes, CEA, List, F-38000 Grenoble, France
[2] Univ. Grenoble Alpes, CEA, Leti, F-38000 Grenoble, France
{firstname.lastname}@cea.fr
[3] Univ. Grenoble Alpes, LPNC, Grenoble, France
{firstname.lastname}@univ-grenoble-alpes.fr

Abstract. While being a key element of human-human communication, face emotion recognition is an important challenge for human-computer interactions. Feature extraction and classification methods have been developed during the past decades in order to propose increasingly accurate emotion recognition algorithms. Nevertheless, in a changing environment where systems needs to be continually adapted, the issue of catastrophic forgetting becomes a major challenge. Based on the bio-inspired continual learning algorithm Dream Net, we propose an embedded system for face emotion recognition. This system is innovative in its ability to learn incrementally on a NVIDIA Jetson Nano platform without catastrophic forgetting while preserving privacy and being agnostic to data. Live demonstration of this system can be done and users can test it in several modes of operation: emotion recognition or learning of new emotions.

Keywords: Facial emotion recognition · Embedded deep learning · Continual learning.

1 Introduction

Emotions are one of the cornerstones of human social interactions. Expressing as well as understanding emotions from others is strongly needed in an environment where several people interact with each other. In today environment where interaction with computers is increasingly common, introducing emotional skills in technologies appears as a way to simplify human-computer interactions [8]. Facial expressions are the most used features in non-verbal communication [7]. Therefore, facial emotion recognition have been particularly studied in past decades and many deep learning methods were proposed [4]. Nevertheless, as they need a large amount of data to be correctly trained, most of the proposed models are not designed to be robust to a changing environment where new emotions or new people can appear. When dealing with human emotion data, especially in the context of facial emotion recognition, privacy becomes an important concern.

Another key issue of deep learning is that "classical" artificial neural network are not able to learn and fine-tune new concepts without a drastic reduction of their performances called "catastrophic forgetting". Based on this observation, Mainsant et al. [6] proposed a bio-inspired continual learning model, called Dream Net, that overcomes catastrophic forgetting, preserves privacy, and is data agnostic. Using this algorithm, we developed an embedded system able to learn and recognize face emotion of people in front of a camera. This work is part of a larger purpose to use the Dream Net algorithm in real life applications such as environmental monitoring, personalization of wearable sensors for healthcare applications or autonomous driving support.

2 Demonstration

2.1 Goal

The system is initially able to recognize five of the seven basic emotions: neutral, angry, disgust, fear and sad. The goal of the demonstration is to extend the system's knowledge to two additional emotions (happy and surprise) without forgetting the five initially learned emotions. This demonstration shows that using Dream Net algorithm allows the system to learn new emotions without storing examples of previous emotions while overcoming catastrophic forgetting.

2.2 Scenario

The demonstration begins with the evaluation of the recognition capabilities of the system. Then, some face images are placed in front of the camera to collect the necessary data for the detector to learn the last two unknown emotions. Next, Dream Net algorithm is used to teach the system these emotions. Finally, the system is updated and tested, not only with face images, but also with real faces. The tests confirm that the system is able to detect all emotions after learning.

2.3 Specifications and Related Optimizations

For the demonstration to run smoothly, the following conditions must be met.First, people seen by the camera should not experience a lag between their movements or emotional changes and the system response: the processing time of the whole system must be less than 100 ms. Second, to keep people's attention, the time it takes for the system to learn new emotions should not exceed one minute.

Such a real-time system has been achieved by using TensorFlow [1] to train the models, and TensorRT [2] to generate 16-bit floating-point optimized runtime engines for inference. Special attention has been paid to minimizing the update time of TensorRT engines once TensorFlow models are trained, and to allow both frameworks to run simultaneously on the same platform, without stepping on each other's toes. A time-memory trade-off was also found due to the small amount of memory available on the embedded platform.

2.4 Performance

The final performance of the emotion detection system is about 10 FPS and the overall learning time for happy and surprise emotions learnt together is about 45 s. There is an increase in emotion detection accuracy of 25% on average compared to a system that does not use a specific continual learning algorithm to overcome catastrophic forgetting.

2.5 Execution

Please see the demonstration video at https://youtu.be/XFVE7vq3iGk

3 Technology

3.1 System Hardware

The system is running on NVIDIA Jetson Nano platform, featuring 128-core GPU and 4GB memory capacity. It is a very popular low-cost embedded platform with great GPU performances, but with a relatively small amount of memory. The platform is enclosed in a metal casing equipped with IMX219-77 camera producing 1280 × 720 pixel resolution images. Finally, a monitor is connected to the HDMI output to display emotion detection results.

3.2 System Architecture

The system pipeline for emotion detection is shown on Figure 1. The face detector is responsible for detecting faces in the camera image. Face images are cropped and resized to grayscale images of 197 × 197 pixels. These images are then fed into the features extractor that outputs feature vectors of size 2048, which are normalized and used by the emotion detector to recognize the associated emotions.

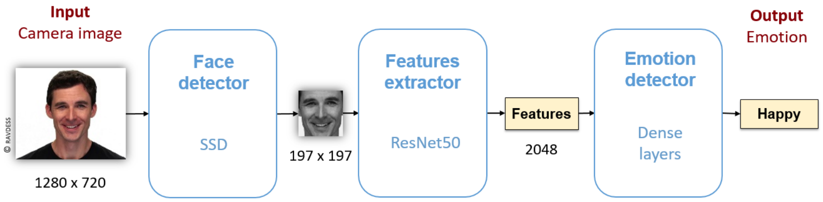

Fig. 1. Pipeline for emotion detection

The face detector model is frozen and provided by OpenCV. It was created with SSD framework [5] using ResNet10 like architecture and trained in Caffe framework. Camera images are scaled to 533 × 300 pixels, knowing that the neural network was initially trained on 300 × 300 pixel images.

The features extractor model is frozen and based on a ResNet50 architecture trained on FER2013 database by Stanford University [3] in which the emotion classifier has been removed. We evaluated the embedded emotion detector trained offline to recognize the seven emotions with this feature extractor and obtained the same accuracy value of 73% on the test set.

The emotion detector model is a hybrid architecture [6] able of replicating the input (like an auto-encoder) and classifying facial emotion in a single inference. It is composed of one input layer of size 2048, one dense layer with 1024 neurons and ReLu activation function, one 50% dropout layer to avoid over-fitting, and one output layer of size 2055 (2048 features replicated and 7 emotions classified) with a sigmoid activation function. This model was trained on FER2013 database where happy and surprise faces have been removed from the training set.

3.3 System Interface

A menu displayed in the execution window allows user to select the desired mode of operation from the three presented below.

The first mode of operation is dedicated to the recognition of live emotions of at most 10 faces simultaneously detected in the camera image. The monitor displays the detected faces enclosed by emotion-annotated bounding boxes.

The second mode of operation is dedicated to generate the "learning dataset" for the continual learning of emotions by the emotion detector. For each emotion to be learnt, several images are captured by the camera, complemented by few images from FER2013 dataset so that the emotion detector can generalize well while learning the new emotions. To preserve privacy, only the associated features are computed and stored in memory.

The third mode of operation is dedicated to learn new emotions. It implements the Dream Net algorithm proposed by Mainsant et al. [6]. The particularity of this model is that it does not store any example of emotions previously learnt because it is able to generate pseudo-examples that represent the past knowledge. New emotions are learnt using these pseudo-examples and the new examples available in the "learning dataset".

4 Conclusion and Future Work

In this paper, we have presented a face emotion recognition system based on the Dream Net algorithm, able to continually learn new emotions without forgetting previous ones. The very good results obtained on NVIDIA Jetson Nano platform demonstrate that Dream Net model can be used on ressource-limted embedded platforms in order to benefit from its two main differentiating properties compared to other continual learning models, namely the agnosticity of the data and the preservation of privacy. Future work will be about bringing personalization and multimodality to the system.

Acknowledgements This demonstrator was developed within the scope of the Carnot MIEL (Multimodal and Incremental Embedded Learning) project. It was also partially supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003).

References

1. TensorFlow framework, `https://www.tensorflow.org`
2. TensorRT framework, `https://developer.nvidia.com/tensorrt`
3. Khanzada, A., Bai, C., Celepcikay, F.T.: Facial Expression Recognition with Deep Learning (2020)
4. Li, S., Deng, W.: Deep Facial Expression Recognition: A Survey (2020)
5. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C.: SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) Computer Vision – ECCV 2016. pp. 21–37. Springer International Publishing, Cham (2016)
6. Mainsant, M., Solinas, M., Reyboz, M., Godin, C., Mermillod, M.: Dream Net: a privacy preserving continual learning model for face emotion recognition. In: 2021 9th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW). IEEE (2021)
7. Revina, I., Emmanuel, W.S.: A Survey on Human Face Expression Recognition Techniques (2018)
8. Wu, C.H., Lin, J.C., Wei, W.L.: Survey on audiovisual emotion recognition: databases, features, and data fusion strategies (2014)