Embedded Deep Learning for Face Detection and Emotion Recognition with Intel® Movidius TM Neural Compute Stick

Anonymous Author(s)
Affiliation
Address
`email`

Abstract

Nowadays deep learning is one of the main topics in almost every field. It helped to get amazing results in a great number of tasks. The main problem is that this kind of learning and consequently neural networks, that can be defined deep, are resource intensive. They need specialized hardware to perform a computation in a reasonable time. Unfortunately, it is not sufficient to make deep learning "usable" in real life. Many tasks are mandatory to be as much as possible real-time. So it is needed to optimize many components such as code, algorithms, numeric accuracy and hardware, to make them "efficient and usable". All these optimizations can help us to produce incredibly accurate and fast learning models.

1 Embedding and face detection

Our work focused on two main tasks that have gained significant attention from researchers, that are automated face detection and emotion recognition. Since these are computationally intensive tasks, not much has been specifically developed or optimized for embedded platforms. We show how inference can be accelerated using Intel's Neural Compute Stick (NCS) [2]. It is a tiny fanless deep learning device, powered by the low power high performance Movidius Myriad 2 Vision Processing Unit (VPU) and allow to accelerate inference optimizing neural networks and operations in order to allow resource-intensive computation on low-resource platforms. We show how this tiny device can let some intensive Deep Learning applications run on embedded devices such as the Raspberry Pi [5] we have used. The inference pipeline is shown in Fig. 1, while Fig. 2 shows how each of the three Neural Compute Stick has been used in the inference acceleration process, two are designed for face detection and the last one to compute the emotion recognition.

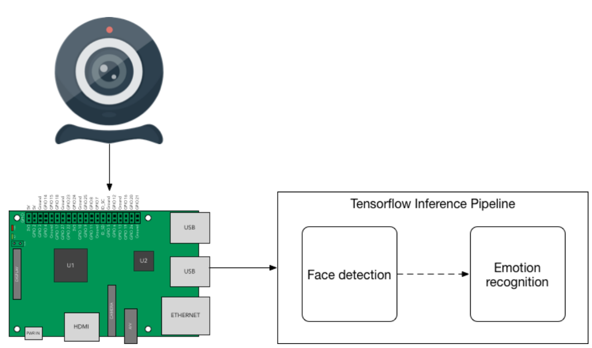


Figure 1: Raspbery pipeline

Submitted to 32nd Conference on Neural Information Processing Systems (NIPS 2018). Do not distribute.

For the first task we have used two of the three networks that characterize a MTCNN [6] and applied a non-maxima-suppression for filtering out outliers in order to get a good trade-off between accuracy and performance.

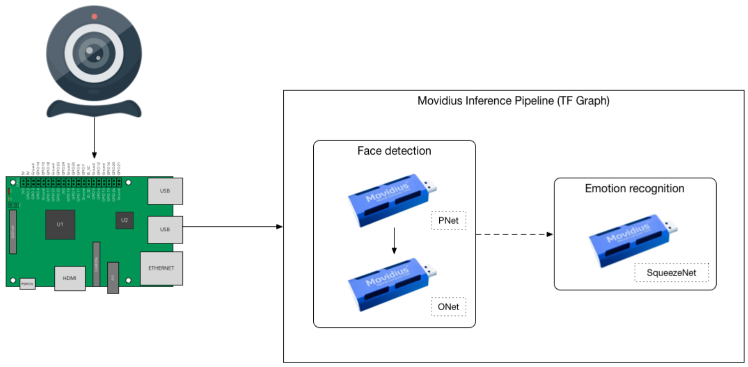
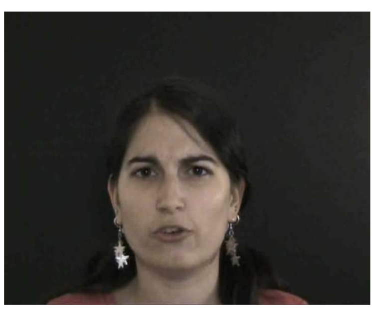


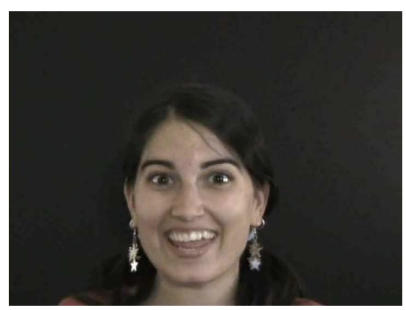
Figure 2: Raspbery + NCSs pipeline

2 Emotion recognition

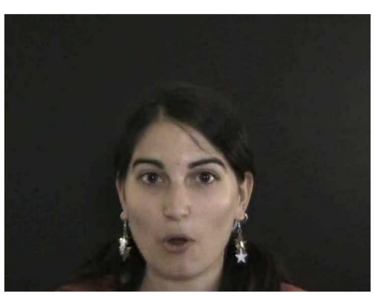
Emotion recognition is a very interesting area to deal with on embedded platforms. It extracts the sentiments starting from face movements. Like many other tasks it needs to be near real-time and it is very difficult to get this kind of performance on embedded devices. We show how the NCS acceleration is able to halve SqueezeNet [1] inference time. This kind of network is used before to reduce the total number of parameters in neural networks like AlexNet [3] and to work efficiently on devices that are not able to handle complex neural networks. Test are done on the eNTERFACE [4] dataset that has 6 classes (Fig. 3).



(a) Anger



(b) Happiness



(c) Surprise

(d) Fear

Figure 3: Example of a subject expressing emotions

3 Conclusions and results

To perform reasonable tests an input image of size 100x100x3 has been used. As shown in Fig. 4 we compared results based on computation time for the pipeline with and without accelerations. Raspberry needs a computation time that is double the time needed by a Movidius, for example the first needs 150ms per frame against the 70ms of the latter. We conducted several tests and reported the inference time for each task and for the whole pipeline in Table 1.

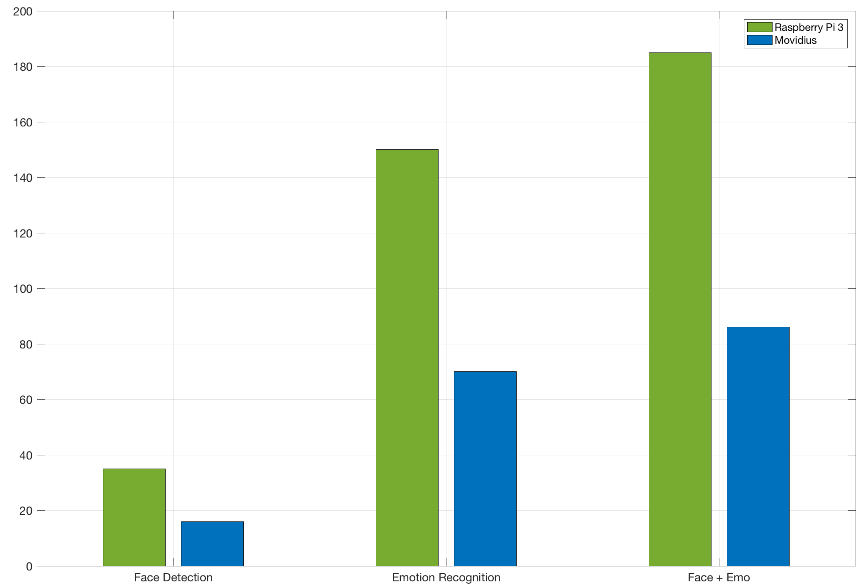


Figure 4: Results comparison

We compared a Tensorflow implementation on a Raspberry Pi 3 B and Intel Movidius NCS. It is possible to state that the NCS is able to speed up the pipeline in order to get results that are almost real-time. Furthermore, it removes some initialization operations that are mandatory with Tensorflow, speeding up also the initial elaboration time.

Device	**Face detection (ms)**	**Emotion recognition (ms)**	**Total + delay (ms)**
Raspberry	35	150	200
Neural Compute Sticks	**16**	**70**	**97**

Table 1: Results comparison for the whole pipeline

References

[1] Forrest N. Iandola, Matthew W. Moskewicz, Khalid Ashraf, Song Han, William J. Dally, and Kurt Keutzer. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size. *CoRR*, abs/1602.07360, 2016.

[2] Intel. Intel® Movidius™ Neural Compute Stick. `https://software.intel.com/en-us/neural-compute-stick`.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[4] Olivier Martin, Irene Kotsia, Benoit Macq, and Ioannis Pitas. The enterface'05 audio-visual emotion database. In *Data Engineering Workshops, 2006. Proceedings. 22nd International Conference on*, pages 8–8. IEEE, 2006.

[5] Raspberry Pi. Raspberry Pi. `https://www.raspberrypi.org/`.

[6] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503, 2016.