Traditional Machine Learning or Deep Learning Methods for Embedded Computer Vision? Study on Biometric Application

Pierre Bonazza, Johel Miteran, Dominique Ginhac and Julien Dubois*

Universite de Bourgogne Franche-Comte, France

Submission: December 01, 2018; **Published:** February 15, 2019

*Corresponding author: Julien Dubois, Université de Bourgogne, Le2i, 9 Avenue Alain Savary, 21000 Dijon, France

Abstract

Today, Convolutional Neural Networks (CNN) provide one of the best performances in state of the art image recognition, while requiring high computational power and being very time consuming. So it raises the following questions: can deep learning achieve better performances than machine learning over different computer vision aspect, such as development time, processing time, raw performances or resource consumption? Considering strong constraints on this listing, can we still consider Machine Learning for a specific computer vision application? Hence, the purpose of this paper is to provide an insight into these questions by comparing methods from both Deep Learning and traditional Machine Learning, applied on a real-time person authentication application, using 2D faces under a binary classification problem. These methods are embedded in a restricted power computational unit, forming a smart camera composing a low-cost security system. This application needs the biometric model to be minimized so it can be stored on a remote personal media (10 KB).

Keywords: Face authentication; CNN; Transfer learning; Machine learning; PCA; SVM; Random forest; Realtime; Embedded security system; Smart camera

Abbreviations: CNN: Convolutional Neural Networks; SVM: Support Vector Machines; RF: Random Forest

Motivations

Designing an embedded system requires taking into account many constraints to make optimal choices, enabling the final system to be profiled considering a software tradeoff. We can thus ask some questions such as" can current low-cost platform make use of deep learning methods?" And if it is the case, is the increasingly popular deep learning now preferable over traditional and mature machine learning in each configuration? Until very recently [1-4], state-of-the-art literature concern were mostly focus on CNNs' raw performances, overlooking target deployment, time and storage aspects. This work aims to give an opinion about these questions by highlighting major differences between those techniques under some strong constraints of a concrete application, namely a biometric security access system using 2D faces. But, setting-up biometric systems poses certain ethical concerns, mostly privacy issues, as some biometric attributes can be acquired unbeknownst to their owner. This incites the possibility of biometry to be used for illegal purposes, which calls for its active supervision. To this end, more and more restrictive legal frameworks try to preserve privacy, recommending the storage of biometric data on individual media, emphasising person authentication rather than person recognition. Consequently, not only in a security context but also in terms of privacy control, the work on which this study focuses is based on data-related aspects to store biometric information on a low storage capacity remote card (10 KB).

Methods and Implementation

This study considers a 2D face authentication, relying on [5,6], using only personal biometrics features per individual, isolated from any network, which have to be stored on a low storage capacity remote card. To make a trade-off between face authentication accuracy and model storage size, meeting real-time and privacy aspects using a multi-core processor embedding all the biometric tasks, multiple classification solutions have been compared on a two-class problem, observing their limitations in terms of algorithm/architecture matching. On the one hand, as shown in Figure 1, classical machine learning methods are employed, needing a feature extraction task before any classification. On the other hand, Convolutional Neural Networks (CNNs) are used, for which all processing steps are merged into a "black box" network. This traditional machine learning feature extraction is achieved using PCA algorithm based on the Eigenfaces method [7], forming a 400 dimensional face space (built with 400 faces from AT&T database [8]) where faces samples are projected.

To reduce the biometric data dimension, PCA thresholds have been intended by truncating eigenvectors (e.vec) to 90% (112 e.vec), 70% (22 e.vec) and 50% (7 e.vec) of the cumulative contribution of their associated eigenvalues. The classification of the feature vectors is achieved using Support Vector Machines

(SVM) and Random Forest (RF). In contrast, transfer learning implementation is put forward by freezing the hidden layers and fine tuning the fully connected layers of a pretrained CNN. The light networks MobileNet v1 [9] and v2 [10] were selected, using their "width multiplier" to respectively 25%* and 35%†. These networks have been originally trained on the object recognition ILSVRC [11], with millions of images spread across 1000 classes.

To further reduce the model storage size [12], traditional machine learning models have been compressed using the BZip2 algorithm (delivering 97,87% of compression on RF and 81,12% on SVM). For an effective compression, CCNs need to be preliminary quantized, and have then been compressed using LZMA algorithm [6] (giving 76% of reduction on those quantized network).
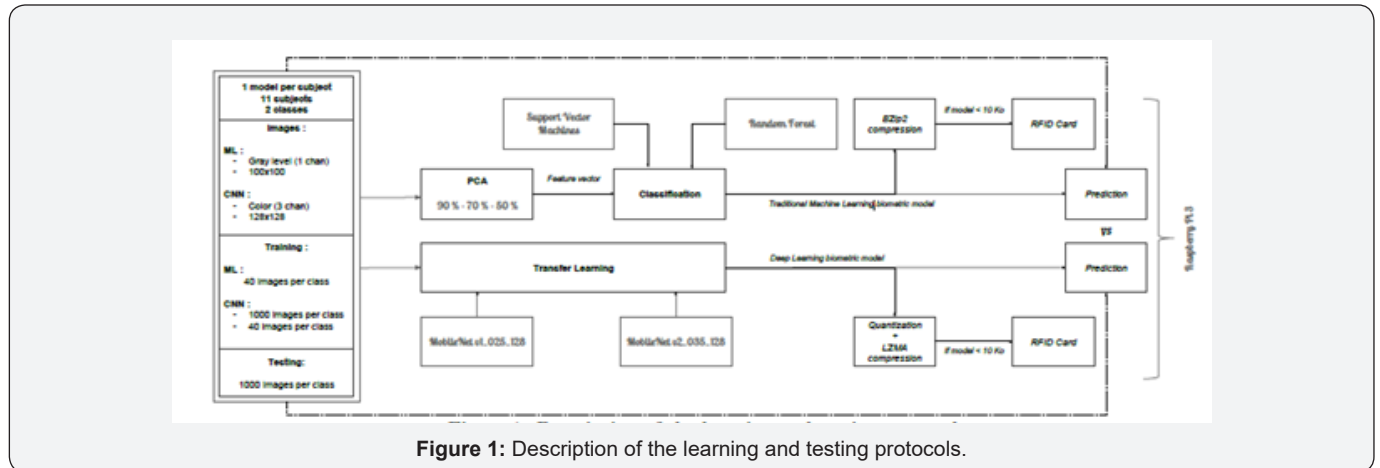

Figure 1: Description of the learning and testing protocols.

Discussion and Conclusion

To summarize the results which are fully presented in [5,6], the SVM achieve between 85% and 94% of authentication accuracy for a model size between 2 KB and 27,3 KB. Its training times fluctuates between 322 ms and 1030 ms with a prediction time for one frame being a minimum of 50 µs and a maximum of 105 µs approximately. Comparatively, RF achieve between 82% and 90% authentication accuracy for a compressed model size between 8,3 KB and 7,9 KB. Its training times fluctuates between 25 ms and 200 ms with a prediction time for one frame about a minimum of 103 µs and a maximum of 117 µs. Considering the framewise PCA projection, the traditional machine learning methods can perform face authentication at 10 fps in average on the target platform. With quantization and compression, the MobileNet v1* and v2† storage sizes are respectively 457 KB and 1625 KB. Using 1000 images per class allows an authentication accuracy of 85% and only 40 images per class slightly decreases the accuracy to 70% on average, as well as the training time, which can reach up to an hour of computation, where the frame prediction is respectively about 1,4 s and 2,1 s for MobileNet v1* and v2†.

Illustrated by a strongly constrained field application, we have tried to highlight crucial elements to determine if deep learning is a better choice than traditional machine learning over various computer vision aspects. We have seen that transfer learning allows to re-train a CNN for a very specific task, raising the accuracy of the lightest quantized MobileNet v1 from 39,5% on ILSVRC (1000 classes), to 85% on face authentication (2 classes) with only 1000 images per class, and 74% for 40 images per class. But the resulting networks are still more time consuming and bulkier than classical machine learning methods which reach 90% of accuracy as well, while being up to 100 times smaller in average. Nevertheless, CNN encompass feature extraction by their hidden layers, which remain the same through all training, in the same way as the PCA. Hence, for a fair comparison, we provide the PCA face space storage size, for respectively the three observed thresholds 90%, 70% and 50%, which are 31,5 MB, 6,4 MB, 2,2 MB, with a training computation time lower than 120 seconds. So we could save only the dense layers on the RFID card and keep the Hidden Layers on the platform, just like the PCA face space. According to [9], the last layers represent 24,33% of all the network parameters. Nonetheless, regarding to the uncompressed network storage size (i.e 1925 KB), the dense layers have to represent less than 0,5% of the total amount of the network parameters to be lower than 10 KB.

Regarding the results, we can say that, for embedded computer vision applications, low computational power platform quickly reaches their limits with the CNNs, which need too many resources to be effective. However, our smartphones are now getting enough computing power to perform this CNN faster than a rapsberry PI 3 [13]. Moreover, in recent times, external computing units are now capable of performing CNN really efficiently. Some of these are in the form of neural sticks, embedding a VPU (Vision Processor Unit), like the "Movidius" Neural Stick [14], delivering 100 GFlop for 1 W, compared to 0,8 GFlops for the same power consumption. Not with standing latest state-of-the-art achievements on CNNs, mature machine learning techniques are still in force when strong constraint problems have to be solved. Without any constraints, and performed on a platform like the Neural Stick, training a CNN from scratch for a dedicated problem will lead to better results.

Acknowledgment

References

1. Iandola FN, Han S, Moskewicz MW, Ashraf K, Dally WJ, et al. (2016) Squeezenet: Alexnet-level accuracy with 50x fewer parameters and¡ 0.5 mb model size. arXiv preprint arXiv:1602.07360.

2. Zhang X, Zhou X, Engxiao Lin, Sun J (2017) Shufflenet: An extremely efficient convolutional neural network for mobile devices. CoRR abs/1707.01083.

3. Wang RJ, Li X, Ao S, Ling CX (2018) Pelee: A real-time object detection system on mobile devices. arXiv preprint arXiv:1804.06882.

4. Freeman I, Roese-Koerner L, Kummert A (2018) Effnet: An efficient structure for convolutional neural networks. arXiv preprint arXiv:1801.06434.

5. Bonazza P, Miteran J, Ginhac D, Dubois J (2018) Comparative study of deep learning and classical methods: smart camera implementation for face authentication, in Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies II. International Society for Optics and Photonics.

6. Bonazza P, Miteran J, Ginhac D, Dubois J (2018) Machine learning vs transfer learning smart camera implementation for face authentication, Proceedings of the 12th ICDSC, Pp. 21.

7. Turk M, Pentland A (1991) Eigenfaces for recognition. Journal of cognitive neuroscience 3(1): 71-86.

8. Samaria FS, Harter AC (1994) Parameterisation of a stochastic model for human face identification. Applications of Computer Vision, Proceedings of the Second IEEE Workshop 138-142.

9. Howard G, Zhu M, Chen B, Kalenichenko D, Wang W, et al. (2017) Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861.

10. Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC (2018) Mobilenetv2: Inverted residuals and linear bottlenecks-mobile networks for classification, detection and segmentation.

11. Deng J, Dong W, Socher R, Li LJ, Li K, et al. (2009) Imagenet: A large-scale hierarchical image database. CVPR 248-255.

12. Bonazza P, Miteran J, Ginhac D, Dubois J (2017) Optimisation conjointe de la taille de stockage et des performances de modeles de classification pour l'authentification de visages. Gretsi.

13. Ignatov A, Timofte R, Szczepaniak P, Chou W, Wang K, et al. (2018) Ai benchmark: Running deep neural networks on android smartphones. arXiv preprint arXiv:1810.01109.

14. Intel movidius neural compute stick, tech. rep.