INCREMENTAL LEARNING WITH PRE-TRAINED CONVOLUTIONAL NEURAL NETWORKS AND BINARY ASSOCIATIVE MEMORIES

Ghouthi Boukli Hacene, Vincent Gripon, Nicolas Farrugia, Mattieu Arzel & Michel Jezequel
Department of Electronics
IMT Atlantique
Brest, France
{name.surname}@imt-atlantique.fr

ABSTRACT

Thanks to their ability to absorb large amounts of data, Convolutional Neural Networks (CNNs) have become the state-of-the-art in various vision challenges, sometimes even on par with biological vision. CNNs rely on optimisation routines that typically require intensive computational power, thus the question of implementing CNNs on embedded architectures is a very active field of research. Of particular interest is the problem of incremental learning, where the device adapts to new observations or classes. To tackle this challenging problem, we propose to combine pre-trained CNNs with Binary Associative Memories, using product random sampling as an intermediate between the two methods. The obtained architecture requires significantly less computational power and memory usage than existing counterparts. Moreover, using various challenging vision datasets we show that the proposed architecture is able to perform one-shot learning – even using only part of the dataset –, while keeping very good accuracy.

1 INTRODUCTION

For the past few years, Deep Neural Networks (DNNs) have achieved state-of-the-art performance (Hong et al. (2015); Pan & Yang (2010); Krizhevsky et al. (2012)) in numerous domains of supervised learning (Cadieu et al. (2014)). CNNs rely on hundreds of millions of parameters that are trained using large amounts of data. In this context, a major drawback of the method is the need for intensive computation and memory during the learning phases. This limitation is critical for embedded systems, such as smartphones or sensor networks.

Incremental methods provide solutions to process the learning data sequentially, using subsets of the training dataset. An incremental technique is defined as such Polikar et al. (2000; 2001): a) it is able to learn additional information from new data (example-incremental), b) it does not require access to the original data used to train the existing classifiers (in order to limit memory usage), c) it preserves previously acquired knowledge (avoid catastrophic forgetting) and d) it is able to accommodate new classes that may be introduced with new data (class-incremental). Most of existing works either add new classifiers to accommodate new data, such as the learn++ method (Polikar et al. (2001); Sun et al. (2016)), or retrain the model using newly received data together with the old model (Syed et al. (1999); Poggio & Cauwenberghs (2001)). To avoid training a large number of classifiers, and to address the *catastrophic forgetting* problem (Kasabov (2013); French (1999)), a combination between SVMs and learn++ method called "SVMlearn++" (Erdem et al. (2005)) was proposed, showing promising improvements (Molina et al. (2014)). However, this method still needs to retrain a new SVM each time new data is provided, and some knowledge is forgotten while new information is being learned.

In this paper we propose an incremental learning model with the following claims: a) it is possible to adapt the model to new data without retaining it, b) the model uses much less computational power than existing counterparts, c) the model approaches state-of-art precision on challenging vision datasets (CIFAR10, ImageNet), d) the model dramatically decreases memory footprint (by

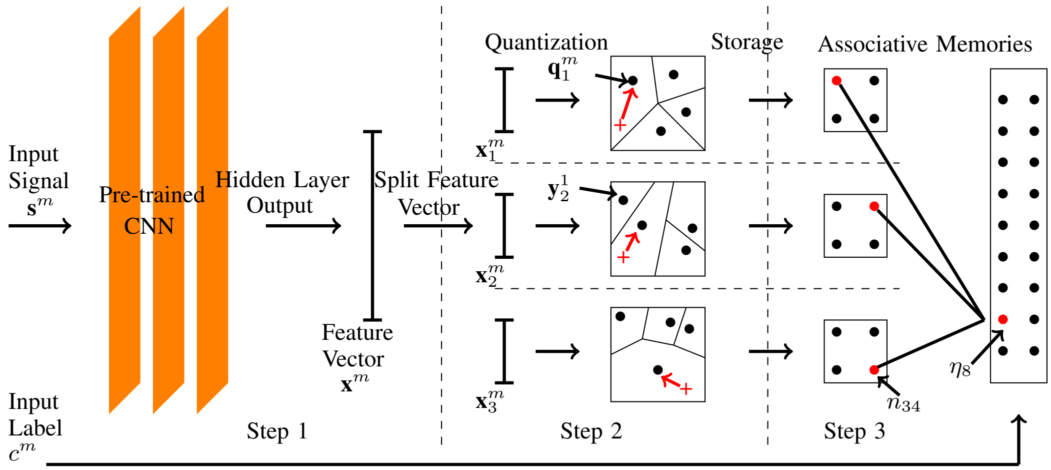

Figure 1: Overview of the proposed method. Our incremental learning method comprises three main steps. Given a set of samples, we first use a pre-trained CNN to extract features (step 1). Subsequently, we use a PQ technique to quantize the feature vectors (step 2). Finally, we use binary associative memories to store and classify the quantized data (step 3).

several orders of magnitude compared to nearest neighbour search), and finally e) the model only requires a few learning examples. An increasingly popular method to benefit from the advantages of DNNs without training is called "transfer learning" (Girshick et al. (2014); Pan & Yang (2010)). The idea is to exploit pre-trained DNNs on large datasets as Feature Extractors (FE), and then fine-tune them using the actual dataset.

We propose to combine transfer learning with binary associative memories. The latter are devices that are able to perform one-shot learning with very limited memory usage and few computational resources. An overview of the proposed method is depicted in Figure 1. We evaluate the proposed method on challenging vision datasets (ImageNet and CIFAR10), and compare both precision and used resources with nonincremental methods.

2 METHOD

The proposed method is built upon three main ideas: 1) using a pre-trained deep CNN to perform features extraction of signals, 2) using product quantizing techniques to embed data in a finite alphabet and 3) using binary associative memories to store and classify data as a proxy to a nearest neighbour search. In the next paragraphs, we detail these three steps.

The first key idea is to use the internal layers of a pre-trained deep CNN (Krizhevsky et al. (2012)) which acts as a generic feature extractor and associates an input signal $\mathbf{s}^m$ with a feature vector $\mathbf{x}^m$ (cf. Figure 1 step 1).

Having a feature vector set $X = \{\mathbf{x}^1, \mathbf{x}^2, \dots, \mathbf{x}^M\}$, the next step is to embed $\mathbf{x}^m$ in a finite alphabet. This step is crucial as it allows to map outputs of step 1) to the inputs of step 3). There is a lot of literature dedicated to this problem, including methods relying on Product Quantization (PQ) (Jegou et al. (2011)). Because we aim at providing computationally light solutions, we rather use product random sampling in this work. Basically, we split each $\mathbf{x}^m$ into P subvectors of equal sizes denoted $\left(\mathbf{x}_p^m\right)_{1\leq p\leq P}$, which are quantized independently from each other using random selection of K anchor points $Y_p = \mathbf{y}_{p1}, ..., \mathbf{y}_{pK}$, where each $\mathbf{y}_p^k$ is such that $\exists \mathbf{x}^m \in X, \mathbf{x}_p^m = \mathbf{y}_p^k$.

After step 2), each feature vector $\mathbf{x}^m$ is transformed into a word of fixed length $\left(\mathbf{q}_p^m\right)_{1\leq p\leq P}$ over a finite alphabet (the alphabet of the anchor points) (cf. Figure 1 step 2). These points are associated with a corresponding output c^m (typically an indicator vector of the class) through a binary sparse associative memory (Gripon & Berrou (2011)). The key idea here is to connect the neurons n_{pk} corresponding to $\mathbf{q}_p^m$, with the neuron $(\eta_c)_{1\leq c\leq C}$ corresponding to the class c^m, where C is the

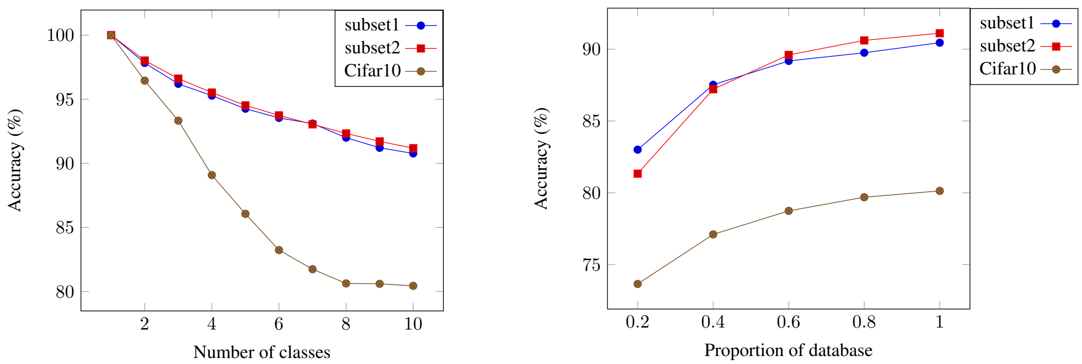

Figure 2: Evolution of the accuracy as a function of number of classes for $P = 16$ and $K = 200$ (ImageNet subset1, subset2 and Cifar10) (left) and as a function of the number of learning examples (right).

number of classes (cf. Figure 1 step 3). We do the same process for every new data allowing incremental learning. Our method is a combination of a deep pre-trained CNN that does not change during the training process, and associative memories that are modified after each newly observed example or class. This combination allows to handle both example and class incremental approaches with no other prior about the learning dataset, using only few learning examples and without having to retrain the model or damage the previously obtained knowledge (Goodfellow et al. (2013)).

3 Results

Results of our simulations on two arbitrarily chosen subsets extracted from ImageNet (containing 10 classes each distinct from the 1000 ones that were used to train the CNN) and Cifar10 are summarised in Figure 2. A comparison in terms of accuracy, computational cost and memory cost with raw Nearest Neighbor (NN) search and accelerated NN search with Product Quantization is summarised in Table 1. As a baseline, we trained a non-incremental linear softmax classifier on the CNN feature vectors. We obtain an accuracy of 89% for Cifar10 and 96.0% for ImageNet subset2.

	Proposed	Other techniques (Preceded by CNN)			
	Method	raw 1 NN	raw 5 NN	PQ 1 NN	PQ 5 NN
Accuracy(%)	82.0/92.0	85.0/95.6	**87.0/96.5**	82.6/94.0	85.0/94.8
Complexity (learning)$\cdot 10^8$	**negligible**	**negligible**	**negligible**	200/49	200/49
Complexity (processing)$\cdot 10^5$	**4.1/4.1**	1000/246	1000/246	32/12	32/12
Memory usage (learning)$\cdot 10^7$	**1.3/1.3**	330/77	330/77	3.7/1.8	3.7/1.8
Memory usage (procession)$\cdot 10^7$	**1.3/1.3**	330/77	330/77	3.7/1.8	3.7/1.8

Table 1: Accuracy, complexity (for learning the whole dataset and for processing one input vector) and memory usage of proposed approach ($P = 64$, $K = 200$) compared to raw NN search and to PQ NN search for Cifar10 / Imagenet subset2.

4 Conclusion

We introduced a novel incremental algorithm based on pre-trained CNNs and associative memories to classify images, the first ones using connection weights to process images, the second one using existence of connections to store them efficiently. This combination of methods allows to learn and process data using very few examples, memory usage and computational power. The obtained precision is close to other state-of-the-art methods based on transfer learning. As a consequence, we believe this method is promising for embedded devices and consider proposing minimal hardware implementations of it as future work.

References

Charles F Cadieu, Ha Hong, Daniel LK Yamins, Nicolas Pinto, Diego Ardila, Ethan A Solomon, Najib J Majaj, and James J DiCarlo. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Comput Biol*, 10(12):e1003963, 2014.

Zeki Erdem, Robi Polikar, Fikret Gurgen, and Nejat Yumusak. Ensemble of svms for incremental learning. In *International Workshop on Multiple Classifier Systems*, pp. 246–256. Springer, 2005.

Robert M French. Catastrophic forgetting in connectionist networks. *Trends in cognitive sciences*, 3(4):128–135, 1999.

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 580–587, 2014.

Ian J Goodfellow, Mehdi Mirza, Da Xiao, Aaron Courville, and Yoshua Bengio. An empirical investigation of catastrophic forgetting in gradient-based neural networks. *arXiv preprint arXiv:1312.6211*, 2013.

Vincent Gripon and Claude Berrou. Sparse neural networks with large learning diversity. *IEEE transactions on neural networks*, 22(7):1087–1096, 2011.

Seunghoon Hong, Tackgeun You, Suha Kwak, and Bohyung Han. Online tracking by learning discriminative saliency map with convolutional neural network. *CoRR*, abs/1502.06796, 2015. URL `http://arxiv.org/abs/1502.06796`.

Herve Jegou, Matthijs Douze, and Cordelia Schmid. Product quantization for nearest neighbor search. *IEEE transactions on pattern analysis and machine intelligence*, 33(1):117–128, 2011.

Nikola Kasabov. *Evolving connectionist systems: Methods and applications in bioinformatics, brain study and intelligent machines*. Springer Science & Business Media, 2013.

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

José Fernando García Molina, Lei Zheng, Metin Sertdemir, Dietmar J Dinter, Stefan Schönberg, and Matthias Rädle. Incremental learning with svm for multimodal classification of prostatic adenocarcinoma. *PloS one*, 9(4):e93600, 2014.

Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

Tomaso Poggio and Gert Cauwenberghs. Incremental and decremental support vector machine learning. *Advances in neural information processing systems*, 13:409, 2001.

Robi Polikar, Lalita Udpa, Satish S Udpa, and Vasant Honavar. Learn++: an incremental learning algorithm for multilayer perceptron networks. In *Acoustics, Speech, and Signal Processing. ICASSP'00. Proceedings.IEEE International Conference on*, volume 6, pp. 3414–3417. IEEE, 2000.

Robi Polikar, Lalita Upda, Satish S Upda, and Vasant Honavar. Learn++: An incremental learning algorithm for supervised neural networks. *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)*, 31(4):497–508, 2001.

Yu Sun, Ke Tang, Leandro L Minku, Shuo Wang, and Xin Yao. Online ensemble learning of data streams with gradually evolved classes. *IEEE Transactions on Knowledge and Data Engineering*, 28(6):1532–1545, 2016.

Nadeem Ahmed Syed, Syed Huan, Liu Kah, and Kay Sung. Incremental learning with support vector machines. 1999.