First exploration of the potential of diverse training and voting for increasing the accuracy of CNNs

Julián Proenza, Yolanda González, Patricia Arguimbau
Departament de Ciencies Matemàtiques i Informàtica
Universitat de les Illes Balears, Spain
julian.proenza@uib.es, yolanda.gonzalez@uib.es, patricia.argimbau@uib.es

This work is supported in part by the Spanish Agencia Estatal de Investigaci´on (AEI) and in part by FEDER funding through grant TEC2015-70313-R (AEI/FEDER, UE).

GOBIERNO DE ESPAÑA MINISTERIO DE ECONOMÍA, INDUSTRIA Y COMPETITIVIDAD
EUROPEAN UNION
EUROPEAN REGIONAL DEVELOPMENT FUND
"A way to make Europe"

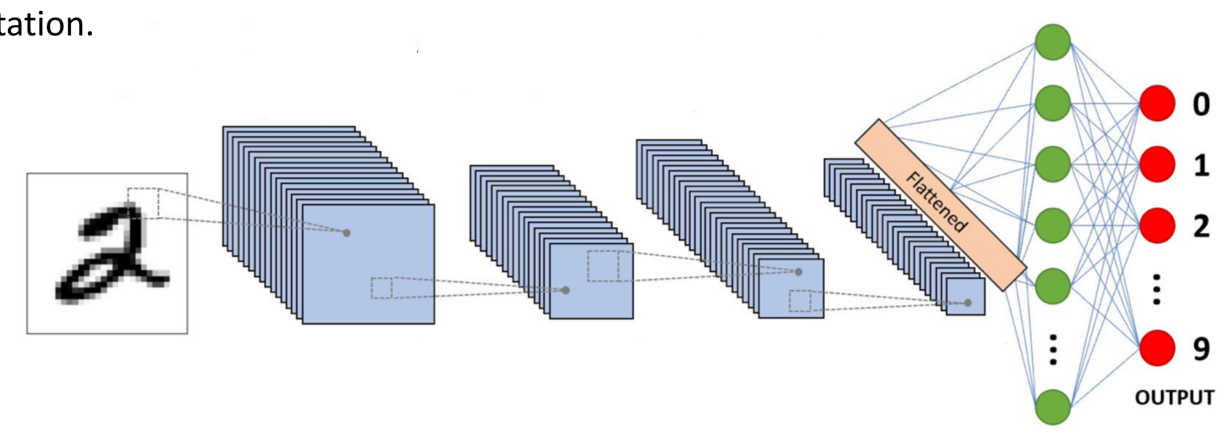
Abstract

Machine learning techniques are attracting a huge amount of interest from both industry and academia. For instance, Convolutional deep Neural Networks (CNNs) have recently enjoyed a notable success in image understanding.
The automotive industry is already using image classifiers for Advanced Driver-Assistance Systems and in the development of the upcoming autonomous cars, which will have to guarantee high levels of reliability. The certification of systems based on machine learning is an open issue but it is clear that any improvement in the performance of image classifiers is to be welcomed. CNNs need to be trained to act as image classifiers.
This training leads to slightly different classification capacity depending on some training parameters.
In this paper we present a first exploration on the use of schemes based on voting on the results of several CNNs trained differently, as a means to the reliability of this type of systems.

1. Convolutional Neural Networks (CNN) for analysing visual imagery

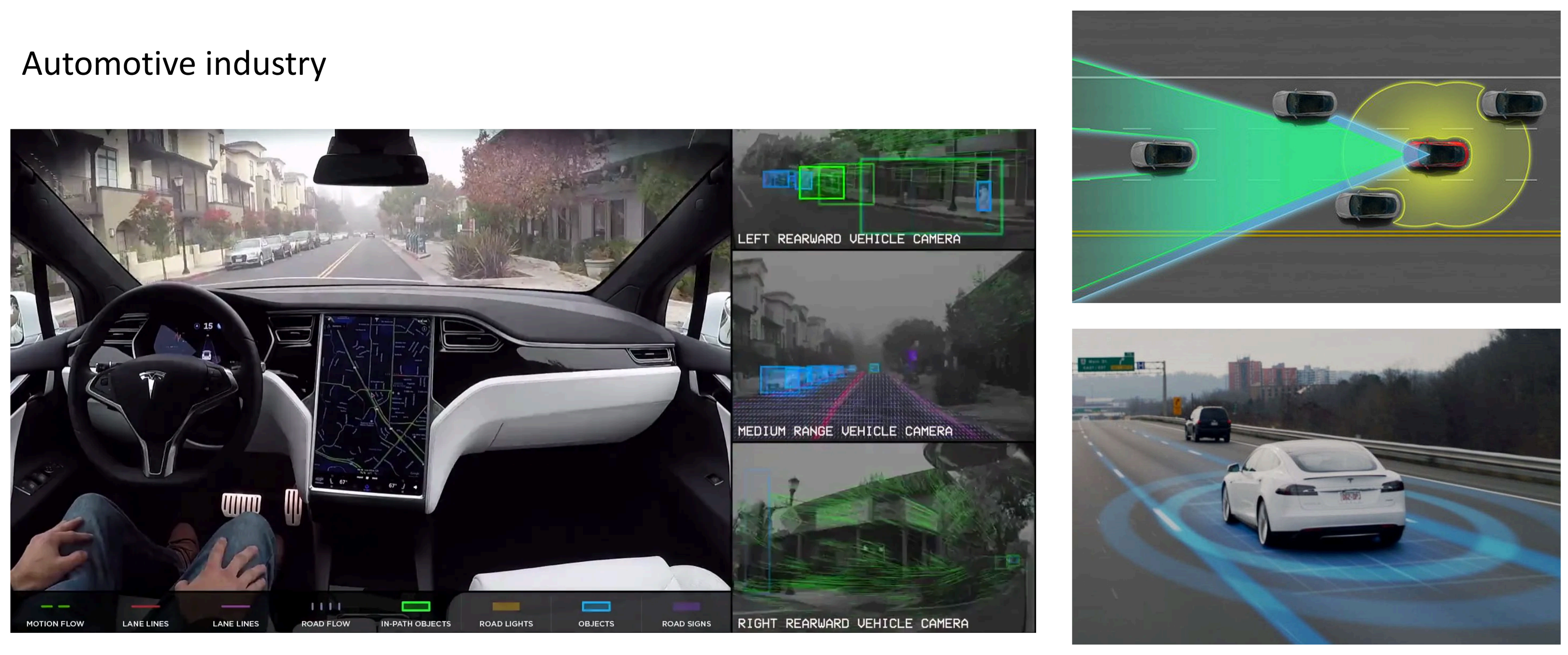
Machine learning is attracting a huge amount of interest from both industry and academia.

Convolutional deep Neural Networks (CNNs) are computing systems that use the machine learning principles and that have recently enjoyed a notable success in image classification or semantic segmentation.



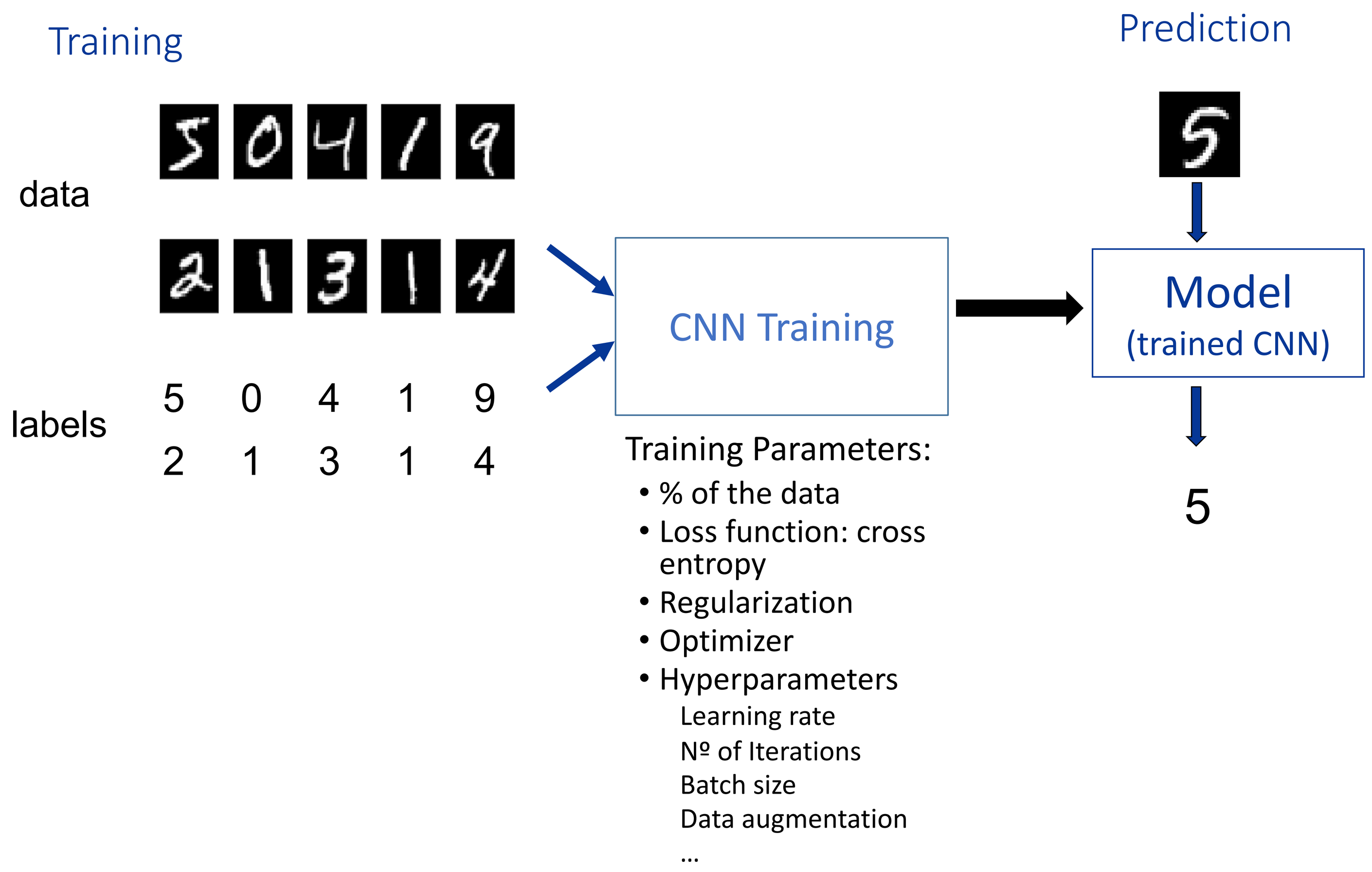
2. Key sectors interested in CNNs

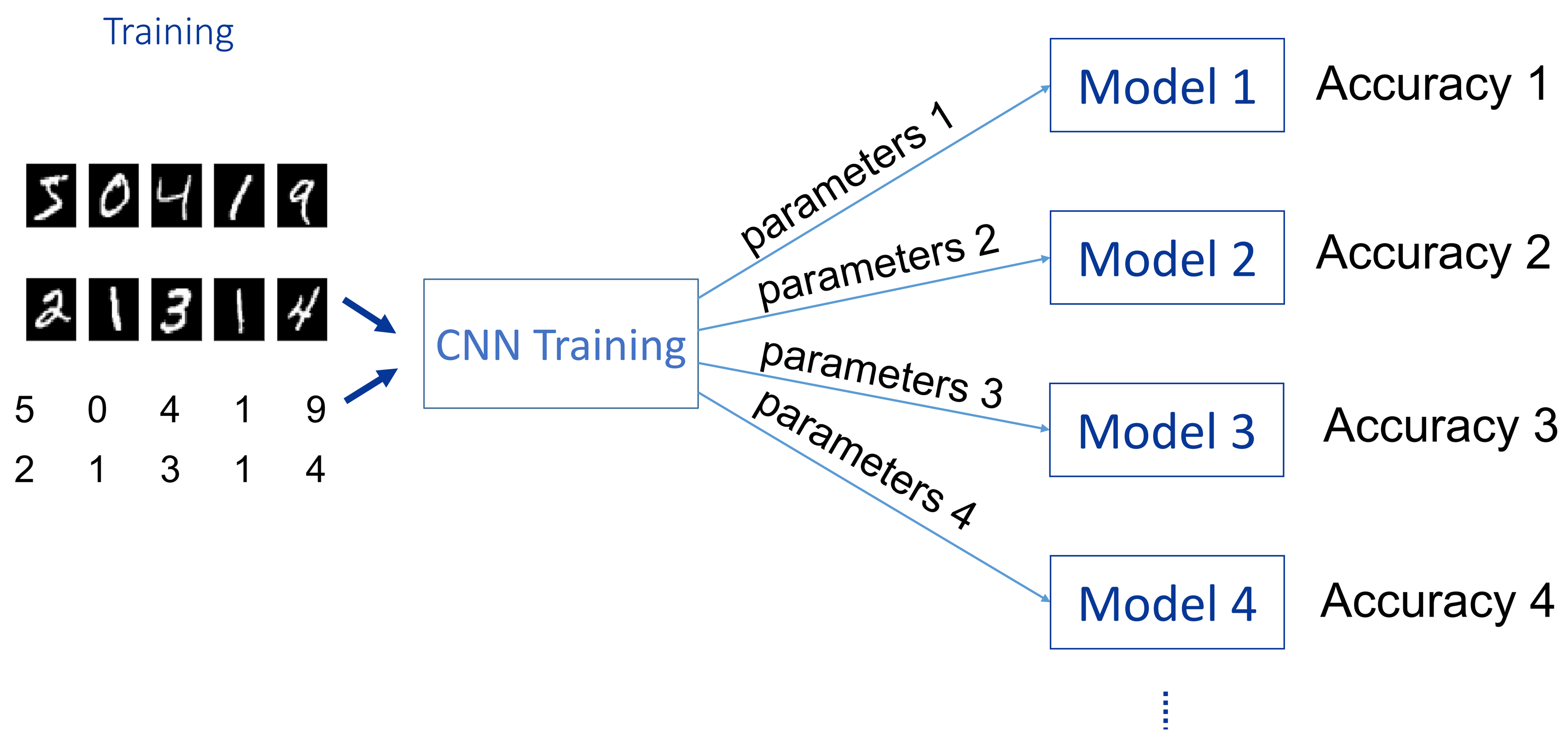
Automotive industry



3. Problem: Diversity in the performance of CNNs

CNNs need to be trained before using them.
This training process leads to slightly different results depending on the **actual set of images** and their **order** in the sequence, among other **hyperparameters** of the CNN architecture.
The same CNN architecture, once trained, could be able to classify a specific test image or not depending on these factors.
To explore this diversity, the authors train a well known CNN using the widely used database of images **MNIST**.

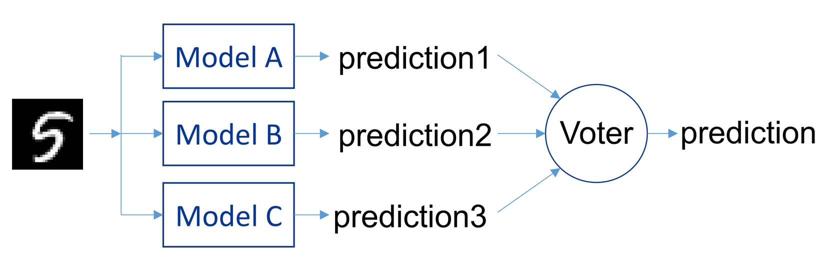




4. Proposed idea: Voting on diverse replicas of a CNN

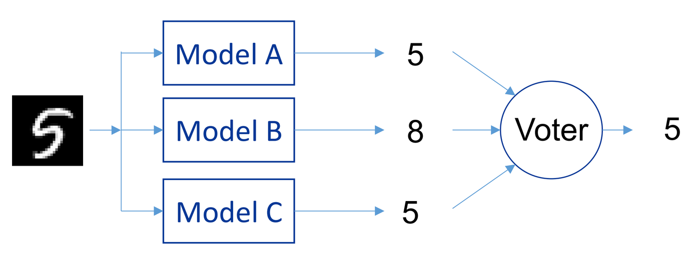
We use N-Modular Redundancy (NMR) and N-version programming (NVP) as basic references, using majority voting.

Let p_i with $i \in \{0,\ldots,9\}$ be the probability (calculated by the CNN) that the input image corresponds to the class of the number i. Considering the case in which there are 3 versions (same CNN architecture trained in three different ways) of the CNN, let ${p_i}^j$ with $j \in \{A,B,C\}$ be the value of p_i as calculated by the j version of the CNN, therefore $P^j = \{{p_i}^j\}_{i=0}^9$ will denote the set of the probabilities of each class calculated by the j version of the CNN .



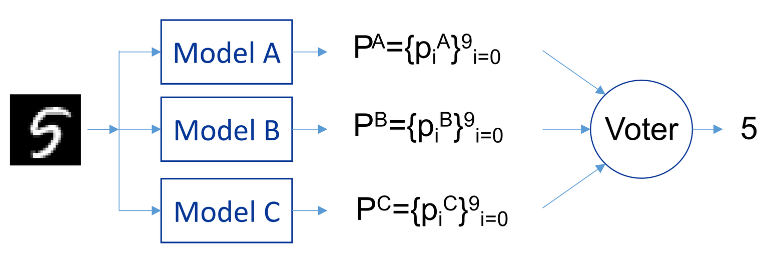
Voting Algorithm 1 (VA1)

Using *Voting Algorithm 1* each version j calculates its own ${p_{\max}}^j = \max P^j$ and it chooses the class corresponding to that ${p_{\max}}^j$ as its output to the voter. Then the voter performs a simple majority voting and only in case two versions propose the same class the voter will be able to produce a, hopefully correct, result.



Voting Algorithm 2 (VA2)

Using *Voting Algorithm 2* each version j only calculates the set of values P^j as output to the voter. The voter then calculates a new set $P^\star = \{{p_i}^A + {p_i}^B + {p_i}^C\}_{i=0}^9$ and then ${p_{\max}}^\star = \max P^\star$. Then it chooses the class corresponding to that ${p_{\max}}^\star$ as final result.



VA2 will potentially outperform VA1 in scenarios in which for the same j there are several p_i^j with values close to the maximum one, but it cannot be anticipated which VA will yield a better accuracy in general, due to the wide variety of scenarios that are possible when considering the potential values of all p_i^j

5. Results

Next table shows the accuracy of simplex and redundant CNN systems for 16 training configurations.

					Simplex Accuracy			VA1 Accuracy			VA2 Accuracy		
Conf.	Number of Versions	Training Images	Batch Size	Loss	Min	Average	Max	Min	Average	Max	Min	Average	Max
1	3	60000	50	0	0.9909	0.992633452	0.9939	0.9927	0.9934	0.994	0.9927	0.99358	0.9941
2	3	60000	50	$< 10^{-6}$	0.987	0.990553459	0.9921						
3	3	60000	100	0	0.9915	0.992286791	0.9936	0.9926	0.99299	0.9936	0.9926	0.9932	0.9939
4	3	60000	100	$< 10^{-6}$	0.9905	0.99177679	0.993						
5	3	20000	50	0	0.9847	0.987453467	0.9891	0.9884	0.98954	0.9902	0.9891	0.99005	0.991
6	3	20000	50	$< 10^{-6}$	0.985	0.986953459	0.9885						
7	3	20000	100	0	0.9844	0.987206803	0.9885	0.9888	0.98938	0.9898	0.9896	0.99009	0.9906
8	3	20000	100	$< 10^{-6}$	0.985	0.986706787	0.9882						
9	5	60000	50	0	0.9913	0.992438127	0.9938	0.9896	0.99357	0.9931	0.9932	0.99369	0.9943
10	5	60000	50	$< 10^{-6}$	0.9852	0.99054413	0.9923						
11	5	60000	100	0	0.9911	0.992240002	0.9936	0.9927	0.99316	0.9936	0.9928	0.99333	0.9936
12	5	60000	100	$< 10^{-6}$	0.9901	0.991709995	0.9935						
13	5	12000	50	0	0.9835	0.985193998	0.9871	0.9864	0.98728	0.9877	0.9872	0.9881	0.9886
14	5	12000	50	$< 10^{-6}$	0.9816	0.984100004	0.9861						
15	5	12000	100	0	0.9821	0.985220129	0.9868	0.9865	0.98763	0.9888	0.9875	0.98831	0.9892
16	5	12000	100	$< 10^{-6}$	0.982	0.984350128	0.9864						

6. Conclusions

There are still too many parameters in the CNNs that are tuned using best-effort approaches. This does not seem the most appropriate way for obtaining the kind of optimal results that are expected when targeting critical applications.

In this context, the **voting techniques** presented in this paper offer the **possibility** of achieving a **better performance in average** without having to put an extra effort in understanding the internals **of the CNN**.

These techniques are **compatible** with any **improvement** that **simplex CNNs** could experience in their performance.

This paper represents a first step towards a better understanding of the potential of diverse training as a means to improve the performance of CNNs.

Universitat de les Illes Balears
systems, robotics & vision
24th INTERNATIONAL CONFERENCE ON EMERGING TECHNOLOGIES AND FACTORY AUTOMATION
ETFA 2019