Integration of Convolutional Neural Networks in Mobile Applications: Platforms and Challenges

1st Roger Creus Castanyer
Universitat Politècnica de Catalunya (UPC)
Barcelona, Spain
roger.creus@estudiantat.upc.edu

Appendix

The results on the CNN models are shown in Table I, where we compare the number of Convolutional Layers, the sizes of the two implemented Fully Connected layers (FC1 and FC2), the total number of parameters, and also the accuracy achieved in the test set with 10 epochs. Furthermore, we write down if the augmented dataset is used, the fixed image sizes chosen and the model weight in MegaBytes (Mb). Note that the input size of FC1 is equal to the number of output features of the convolutional blocks. Also note that the output size of FC1 is the input size of FC2 and the output size of FC2 is 43, which is the number of unique classes of traffic signs in the data. The resulting optimal configuration is promising because of the little complexity, the small storage weight and the great performance shown.

For the CNN configurations, a part from all the parameters that are tuned there have been set fixed hyperparameters for the models executions. They are listed in Table II.

As it can be seen in Table I, no increase in performance appears when the augmented set of data is used. This might be because very fast training is happening, thanks to the non-saturating ReLU activation function in combination of the convolutional operations plus the use of minibatches when training. More importantly, it can be seen that the highest tier classification accuracy can be achieved with the lightest of the tested configurations.

Regarding the MBv2, the experiments do not change its architecture but only its training methodology. Three techniques are distinguished: Feature Extraction (FE), which uses a pre-trained version of the model and just tunes the built classifier on top of it, training its weights and biases. Fine-Tuning (FT), which also uses the pre-trained version of the model but tunes the weights of all its layers. And Training from Scratch (TfS), which just loads the architecture with a random initialization and fully trains it. The results on the training stage with the MBv2 for 10 epochs are presented in Table III.

An important drawback of the MBv2 is spotted in Table III, where it can be seen that without the augmented set of data the network is not capable of achieving learning on the traffic sign specific task. Anyway, when the augmented set of data is used it can be seen that the same performance as the best CNN can be achieved with a much smaller number of parameters, proving that this architecture is efficient and works as expected.

Finally, the results on the benchmark RN34 with transferred learning is shown in Table IV. Training experiments follow the same logic as with the MBv2 but without trying the augmented set of data, since it rapidly seen that it is not needed for this architecture.

As it is seen in Table IV the training method of FT gets to the highest seen accuracy in the test set during the practical study. The output model of this training stage is the one deployed in the application.

Data Aug.	**Image Size**	**Batch Size**	**Conv. Layers**	**FC1 input**	**FC2 input**	**Num. Parameters**	**Test accuracy**	**Model Weight**
NO	256x256	256	5	8x8x256	4x4x64	17.21M	88.02%	65.68Mb
NO	512x512	64	6	8x8x512	4x4x128	67.13M	95%	259.7Mb
YES	256x256	64	5	8x8x256	4x4x64	17.21M	89.31%	65.68Mb
YES	512x512	64	6	8x8x512	4x4x128	67.13M	**95.52%**	259.7Mb
NO	256x256	64	6	4x4x512	2x2x128	**5.1M**	**95%**	**22.11Mb**

TABLE I: Performance analysis of the different CNN configurations.

Conv. Dropout	**FC Dropout**	**Conv. Kernel Size**	**Activation**	**Optimizer**	**Learning Rate**	**Learning Rate Scheduler**	**Max Pool Kernel**	**Weight initialization**
0.2	0.4	3x3	ReLU	SGD(momentum=0.9)	0.01	StepLR(steps=3)	2x2	Xavier's normal

TABLE II: Fixed hyperparameters in the CNN models.

Data Aug.	**Training method**	**Batch Size**	**Test accuracy**
NO	FE	256	4%
NO	FT	256	4%
NO	TfS	256	4%
YES	FE	1024	68%
YES	FT	256	**95.2%**
YES	TfS	256	92.5%

TABLE III: Performance analysis of the Mobile Net v2

Data Aug.	**Training method**	**Batch Size**	**Test accuracy**
NO	FE	256	20%
NO	FT	256	**97.8%**
NO	TfS	256	94.1%

TABLE IV: Performance analysis of the RN34