Traffic Signs Classification using
Convolutional Spiking Neural Networks

Dominika Przewlocka-Rus & Tomasz Kryjak

Embedded Vision Group, Computer Vision Laboratory,

AGH University of Science and Technology, Krakow, Poland

{dominika.przewlocka, tomasz.kryjak}@agh.edu.pl

Abstract: In recent years, neural networks, and especially deep convolutional neural networks, have practically dominated the field of image classification. A specific and less popular subgroup of artificial neural networks are Spiking Neural Networks (SNNs). Unlike classical ones, they reflect the behaviour of biological neurons much more closely. Two basic properties of impulse models are: taking into account the time dimension and event-based operation. The first should potentially translate into greater computational abilities. The second means that the computations are not synchronised with an external clock (synchronous), as in standard processors, but are only executed when needed (asynchronous). Therefore, spiking networks running on dedicated neuromorphic platforms, such as Intel Loihi, can operate with greater energy efficiency, hence they are an interesting alternative for embedded solutions. Unfortunately, their simulation on general-purpose computers, graphics processors, or even FPGAs is not very effective, because all the mentioned platforms are synchronous. To explore the possibilities of using spiking neural networks for embedded vision systems, we have focused on the traffic signs classification problem. We have proposed a convolutional spiking neural network based on the LeNet5 architecture, with a Leaky-Integrate-and-Fire neuron model. For design and simulation, we have used the Nengo and NengoDL libraries for Python language. We have achieved up to 97% accuracy on the test set, which is comparable to static convolutional neural networks with similar architectures.

Work presented in this poster was supported by the National Science Center project no. 2016/23/D/ST6/01389

Training Data

For the classification of traffic signs, the GTSRB (German Traffic Sign Recognition Benchmark) database was used [1]. It consists of images of German traffic signs representing 43 classes. 12630 images are used as the test set and 39209 as the training set, however the number of training samples belonging to each class is not balanced. Each image is in the RGB colour space of size varying between 15×15 up to 250×250 pixels. At the same time, not all are square and differ in illumination and quality. Keeping the above, the images had to be standardised before training.

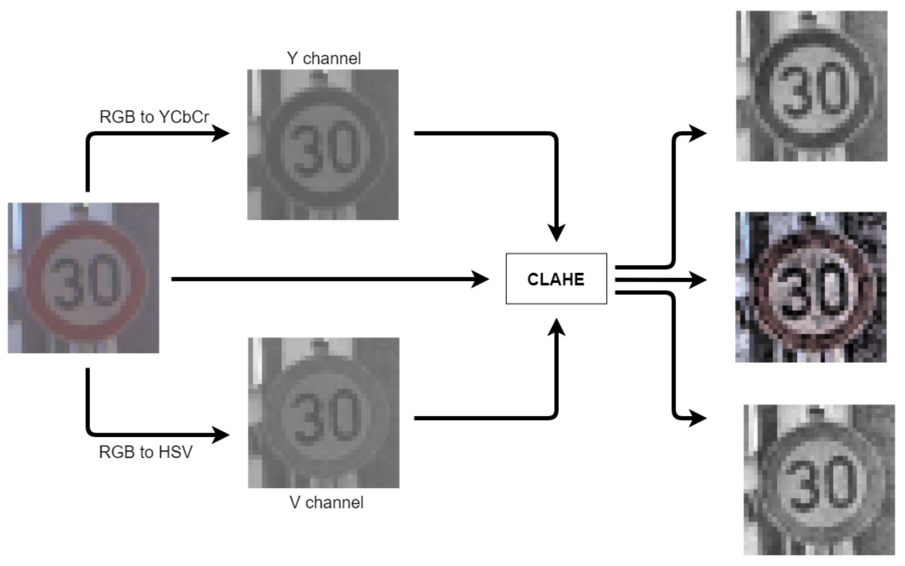


Figure 1: Comparison of different preprocessing methods.

The preprocessing of the input data was split into two parts: (1) data augmentation to balance class representations and (2) standardisation of images. Three types of standardisation were proposed based on CLAHE (Contrast Limited Adaptive Histogram Equalisation) for greyscale (Y and V channels from respectively the YCbCr and HSV colour spaces) and colour images. After applying the corresponding colour space transformation, the image size was normalised to 32x32 pixels. The used preprocessing variants are presented in Figure 1.

Data augmentation was performed randomly for each training experiment using rotation, translation, shear and scaling operations. Several experiments have been carried out with different augmentation methods: equalising class representations or increasing the number of poorly represented classes. Apart from the histogram equalisation, the input data was also centred by subtracting the mean value of the training set (for each channel separately, if applicable). For baseline experiments (static convolutional neural networks for comparison), the input data after mean subtraction was also divided by the standard deviation, as it significantly increased the networks ability to learn.

Spiking Neural Network Architecture

Layer	Input	Neurons	Stride
CONV 1	$32 \times 32 \times 1$	$32 \times [3 \times 3 \times 1]$	$(2, 2)$
CONV 2	$15 \times 15 \times 32$	$64 \times [3 \times 3 \times 32]$	$(2, 2)$
CONV 3	$7 \times 7 \times 64$	$128 \times [3 \times 3 \times 64]$	$(1, 1)$
FC 1	$5 \times 5 \times 64$	120	-
FC 2	120	84	-
FC 3	84	43	-

Table 1: The proposed SNN architecture.

For the traffic sign classification problem, we have proposed a network architecture based on LeNet5 – for subsequent convolutional layers 32, 64 and 128 filters were used. Instead of a pooling layer, we have defined our convolution in the 1st and 2nd convolutional layer with stride equal to 2. However, a number of preliminary experiments were performed with larger and smaller architectures, with varying numbers of layers and filters. On their basis, the architecture most suited to the problem was pre-selected, i.e. the smallest, by means of which good classification results were obtained. The proposed architecture is summarised in Table 1 (input values for greyscale image – for colour 3 input and filters channels in the first layer).

Model Evaluation

To evaluate convolutional spiking neural networks in the traffic sign classification task multiple experiments were conducted – their results are summarised in Table 2. They differ mainly in the input data preprocessing, but also in the data augmentation. Partial augmentation stands for augmenting only the classes with the lowest representation – in our experiments, less than 1000 examples in the training set. Equalised augmentation means that each class is represented by the same number of samples. The training dropout value is also shown, both for convolutional (CONV) and dense (FC) layers. For spiking neural networks, the number of time steps for which the input image is presented is equally important. Moreover, along with each networks accuracy, accuracy of the analogous static network – baseline – with ReLU activation is indicated (in brackets).

	Preprocessing	Augmentation	Dropout (CONV—FC)	Steps	Accuracy
1a	Grey (Y)	Partial (150%)	0.3 — 0.2	30	96.56% (97.84%)
1b	Grey (Y)	Equal (3000)	0.3 — 0.2	30	96.32% (97.13%)
2a	RGB	Partial (150%)	0.3 — 0.2	30	94.94% (96.59%)
2b	RGB	Equal (3000)	0.3 — 0.2	30	94.82% (95.64%)
3	Grey (V)	Partial (150%)	0.3 — 0.2	30	93.03% (95.2%)

Table 2: Different training experiments.

Perhaps the most interesting conclusion from Table 2 should be the fact that the difference in accuracy between the spiking and static network is around 1% for each experiment. To further investigate this differences, several additional experiments on the impact of the number of time steps the input image is presented to SNN were performed. For this and further analysis, the network from 1a experiment is chosen (cf. Tab. 2). The results are shown in Table 3.

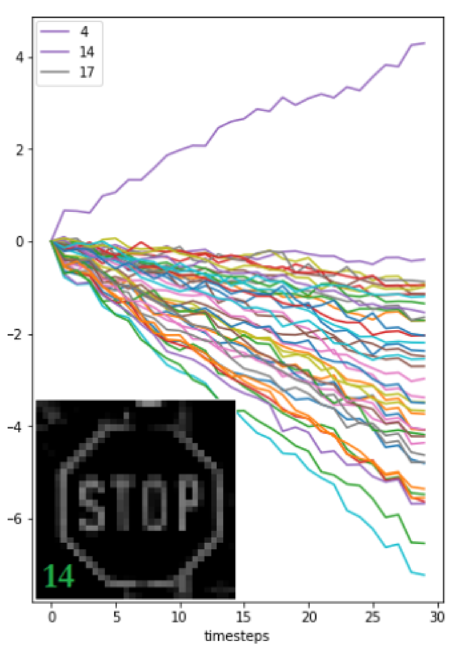
Steps	10	20	30	40	50	100
Accuracy	94.46%	96.07%	96.56%	96.57%	96.66%	97.00%

Table 3: The impact of the number of time steps the input image is presented to the spiking neural network on the accuracy of the test set. Experiments performed for SNN 1a.

The longer the input images are presented to the spiking network, the better the accuracy. This is rather not surprising, however the consequences are important to acknowledge – the increased presenting time results in greater processing time.

Conclusions

An exemplar classification output is presented below – the plot represents the response of the output neurons layer (Y-axis) over time (X-axis) – the highest positive value indicates the classification result.



For good quality and distinguishable traffic signs, such as STOP (class 14), the network is able to predict with great certainty the accurate class (almost from the first time step). What is more, while comparing our work to other classifiers, particularly to the ones presented at the IJCNN [1] challenge, we place ourselves in 3rd place, slightly below human performance (98.84%).

However, demonstrating the practical usefulness of such a solution remains an important problem. In theory, such a network, running on dedicated neuromorphic hardware, such as Intel's Loihi, should be characterised by low latency and low energy consumption. So in order to answer the questions whether and how much more effective is the spiking version than the static one it seems advisable to launch the solution on the Loihi platform – which is planned for a future work.

References

[1] Johannes Stallkamp, Marc Schlipsing, Jan Salmen, and Christian Igel. The German Traffic Sign Recognition Benchmark: A multi-class classification competition. In *IEEE International Joint Conference on Neural Networks*, pages 1453–1460, 2011.