Multiple-dataset traffic sign classification with OneCNN[1]

Fran Jurišić, Ivan Filković, Zoran Kalafatić

University of Zagreb
Faculty of Electrical Engineering and Computing
Department of Electronics, Microelectronics, Computer and Intelligent Systems

ivan.filkovic@fer.hr, fran.jurisic@fer.hr, zoran.kalafatic@fer.hr

22.09.2015.

[1] Fran Jurišić, Ivan Filković, and Zoran Kalafatić. "Multi-dataset traffic sign classification with OneCNN". . In: *IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia*. 2015.

Content

Motivation

Applications

- Advance Driver Assistance Systems - ADAS
- Systems for mapping and assessing the state of traffic infrastructure
- Autonomous vehicles

Existing commercial systems

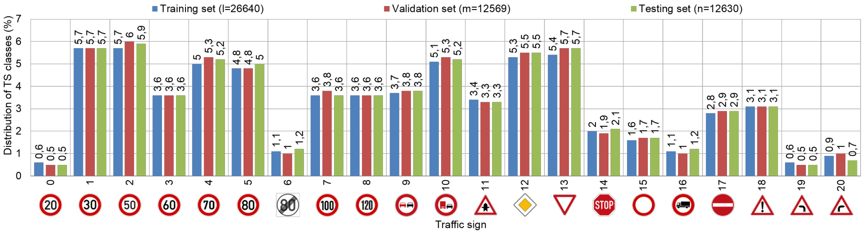
- Possibility of improving performance
- Removing restriction like driving in non-urban areas (e.g. only on motorways)
- Increase detectable and recognizable subset of traffic signs (supported only speed limit signs)

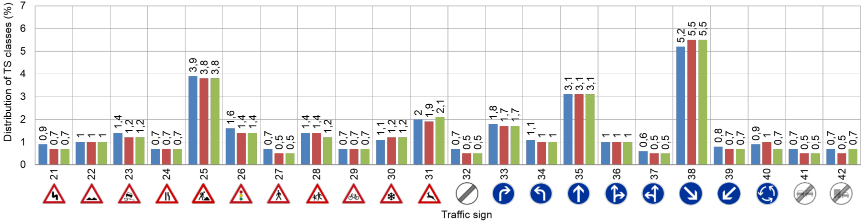
Traffic Signs

Problems

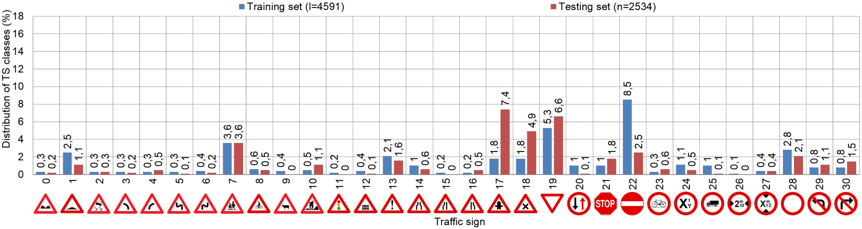
- Large number of traffic signs categories
- Vast amount of good quality examples is required
- Changing lighting and weather conditions
- Partial occlusion
- Different perspectives
- Motion-blur
- Physical degradation, stickers and graffiti
- Visually similar subsets
- Standardization

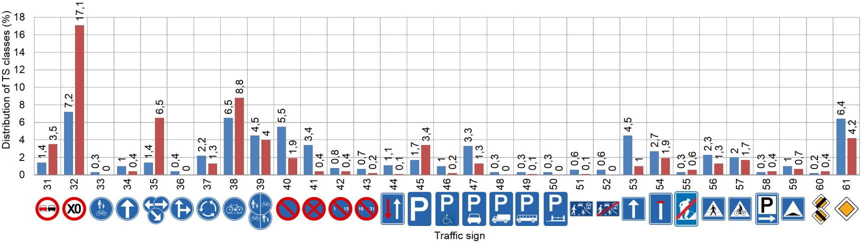
Traffic Sign Classification Datasets - GTSRB





Traffic Sign Classification Datasets - BTSC



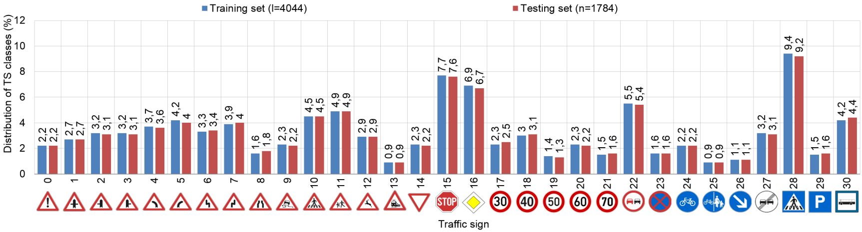


Traffic Sign Classification Datasets - rMASTIF

- New dataset of Croatian signs

Motivation

- Correct possible shortcomings recognized from GTSRB and BTSC datasets
- Addressing traffic signs standardization problem with cross-dataset experiments



Convolutional Neural Networks

- Inspired by simple and complex cells in the primary visual cortex of a brain
- Differ in training procedure and in implementation of convolutional and sub-sampling layers
- Convolutional layers
  - Number of maps (M)
  - Size of maps ($m_{x,y}$)
  - Kernel sizes ($k_{x,y}$)
  - Skipping factors ($s_{x,y}$)
- Sub-sampling layers
  - Max-pooling
  - Average-pooling
- Output
  - Downscaling the output maps of the last convolutional layer to 1 pixel per map
  - Combining the output of last convolutional layer into feature vector

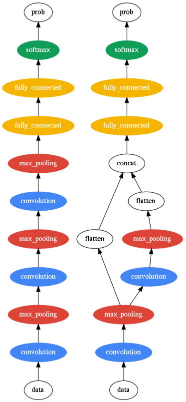
OneCNN

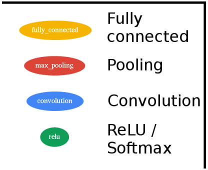
- Based on Ciresan *et al.* [1] (committee of CNNs) and Sermanet *et al.* [6] (Multi-scale CNN)

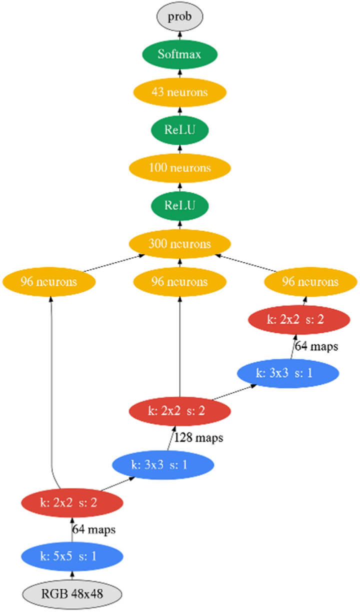
Difference

- More deeper and complex network, but a single one in model
- Added one fully connected layer on top of each convolutional layer
  $\Rightarrow$ Extracted smaller number of abstract features on a per-scale level
- Dropout

OneCNN


prob
softmax
fully_connected
fully_connected
max_pooling
convolution
max_pooling
convolution
max_pooling
convolution
data
prob
softmax
fully_connected
fully_connected
concat
flatten
flatten
max_pooling
convolution
max_pooling
convolution
data


fully_connected
Fully connected
max_pooling
Pooling
convolution
Convolution
relu
ReLU / Softmax


prob
Softmax
43 neurons
ReLU
100 neurons
ReLU
300 neurons
96 neurons
96 neurons
96 neurons
k: 2x2 s: 2
64 maps
k: 3x3 s: 1
k: 2x2 s: 2
128 maps
k: 3x3 s: 1
k: 2x2 s: 2
64 maps
k: 5x5 s: 1
RGB 48x48

Dataset Results

Paper	method	CCR (%)
GTSRB final phase test set		
Jin *et al.* [3]	HLSGD (20×CNN; ensemble)	99.65
Zhu *et al.* [8]	SVM+LLC+SPM(SIFT,HOG,LBP)	99.64 ± 0.018
Ciresan *et al.* [2, 1]	MCDNN (25×CNN; committee)	99.46
Stallkamp *et al.* [7]	Human (best)	99.22
Ours	**OneCNN (1×CNN)**	$\mathbf{99.11 \pm 0.10}$
Jin *et al.* [3]	HLSGD (1×CNN)	98.96 ± 0.20
Stallkamp *et al.* [7]	Human (avg)	98.84
Mathias *et al.* [5]	INNC+INNLP(I,PI,HOGs)	98.53
Ciresan *et al.* [2, 1]	MCDNN (1×CNN)	98.52 ± 0.15
Sermanet *et al.* [6, 7]	MSCNN	98.31
BTSC		
Zhu *et al.* [8]	SVM+LLC+SPM(SIFT,HOG,LBP)	98.77
Mathias *et al.* [5]	INNC+SLRP(I,PI,HOGs)	98.32
Ours	**OneCNN (1×CNN)**	$\mathbf{98.17 \pm 0.22}$
rMASTIF		
Ours	**OneCNN (1×CNN)**	$\mathbf{99.53 \pm 0.10}$

Table: Correct classification rates for different methods.

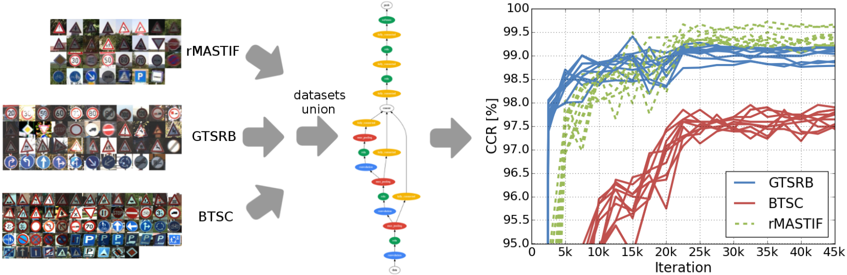
Cross-Dataset Results

train/test	GTSRB	BTSC	rMASTIF
GTSRB	99.11 ± 0.10	98.34 ± 0.17	99.53 ± 0.75
BTSC	86.40 ± 0.61	98.17 ± 0.22	75.37 ± 0.86
rMASTIF	91.06 ± 0.57	97.71 ± 0.47	99.53 ± 0.10
Union	99.04 ± 0.14	97.66 ± 0.16	99.37 ± 0.16

Table: Correct classification rate for different combinations of datasets. The rows stand for training sets, and columns for test sets.

- $|GTSRB \cap BTSC| = 20, |GTSRB \cap rMASTIF| = 18, |BTSC \cap rMASTIF| = 15$
- Asymmetrical results $\Rightarrow$ visual disparities between datasets our CNN failed to bridge
- Union dataset with 87 classes $\Rightarrow$ possible European Traffic Sign Classification dataset

Cross-Dataset Results



Conclusion and Future Goals

- New dataset of Croatian signs
- Single CNN model
- Evaluated performance on multiple datasets, their intersection and union
  $\Rightarrow$ Surpassed known human and other single CNN performance
- Organizing existing datasets
- New performance measures closer to real world application

References I

Dan C. Ciresan, Ueli Meier, and Jürgen Schmidhuber. "Multi-column deep neural networks for image classification." In: *CVPR*. IEEE, 2012, pp. 3642–3649.

Dan C. Ciresan et al. "Multi-column deep neural network for traffic sign classification". In: *Neural Networks* 32 (2012), pp. 333–338.

Junqi Jin, Kun Fu, and Changshui Zhang. "Traffic Sign Recognition With Hinge Loss Trained Convolutional Neural Networks". In: *Intelligent Transportation Systems, IEEE Transactions on* PP.99 (2014), pp. 1–10.

Fran Jurišić, Ivan Filković, and Zoran Kalafatić. "Multi-dataset traffic sign classification with OneCNN". In: *IAPR Asian Conference on Pattern Recognition (ACPR), Kuala Lumpur, Malaysia*. 2015.

References II

- Markus Mathias et al. "Traffic sign recognition - How far are we from the solution?" In: *IJCNN*. IEEE, 2013, pp. 1–8.
- Pierre Sermanet and Yann LeCun. "Traffic sign recognition with multi-scale Convolutional Networks." In: *IJCNN*. IEEE, 2011, pp. 2809–2813.
- Johannes Stallkamp et al. "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition". In: *Neural Networks* 32 (2012), pp. 323–332.
- Yingying Zhu et al. "Traffic sign classification using two-layer image representation". In: *IEEE International Conference on Image Processing, ICIP 2013, Melbourne, Australia, September 15-18, 2013*. 2013, pp. 3755–3759.

The End