Traffic Sign Recognition with Neural Networks in the Frequency Domain

Florian Franzen¹ & Chunrong Yuan¹ & Zhong Li²

¹ Autonomous Systems Lab, TH Köln - University of Applied Sciences, Betzdorfer Str. 2, 50679 Cologne, Germany
² FernUniversität in Hagen, Faculty of Mathematics and Computer Science, Universitätstr. 47, 58097 Hagen, Germany
E-mail: florian.franzen@th-koeln.de

Abstract. In this paper we describe traffic sign recognition with neural networks in the frequency domain. Traffic signs exist in all countries to regulate the traffic of vehicles and pedestrians. Each country has its own set of traffic signs that are more or less similar. They consist of a set of abstract forms, symbols, numbers and letters, which are combined into different signs. Automatic traffic sign recognition is important for driver assistance systems and for autonomous driving. Traffic sign recognition is a subtype of image recognition. The traffic signs are usually recorded by a camera and must be recognized in real time, i.e. assigned to a class. We use neural networks for traffic sign recognition. The special feature of our method is that the traffic sign recognition does not take place in the spatial domain but in the frequency domain. This has advantages because it is possible to significantly reduce the number of neurons and thus the computing effort of the neural network compared to a conventional neural network.

1. Introduction
Traffic signs regulate traffic in all countries. In order to avoid confusion and improve recognition across countries, efforts are being made to harmonise the appearance of traffic signs. This is the aim of the Vienna Convention on Road Signs and Signals, which was developed in 1968 and ratified by numerous countries. With the introduction of Advanced Driver Assistance Systems and the goal of autonomous driving, automatic traffic sign recognition (TSR) has become more important. Video cameras are installed in vehicles for this purpose. The images are continuously recorded and evaluated.

The recognition of traffic signs is done in two steps: In the first step, the location of the traffic sign is detected and the traffic sign is extracted. The classical detection methods can be divided into color based and shape based [1][2]. Based on the camera image of the captured environment, a bounding box can be found which encloses the detected traffic sign. In a second step, the traffic sign is recognized, i.e. assigned to a class. A variety of methods are also possible. In [3] an image database is created with each traffic sign in optimal symbol representation. To recognize an unknown traffic sign, its image is compared with each image from the database and the similarity is calculated using a cross correlation matrix. In a study by [4] the contours of symbols in traffic signs are extracted and described by Fourier descriptors. The assignment to a class is done without neural networks only using the correlation of the Fourier descriptors. It is also possible to extract features such as Histogram of Oriented Gradients or Color Histogramms.
from traffic sign images and classify them using Linear Discriminant Analysis methods [5]. The use of Random Forests is also possible [6].

Since the invention of Convolutional Neural Networks (CNN) by Yann LeCun [7], most programs for image recognition use CNN. For example, they are successfully used in [8][9][10] for TSR. Similar to wavelet transform, the extracted features represent location and frequency information at one time [11]. In our approach we go one step further: We convert the images purely into the frequency domain and then have the classification carried out by a neural network [12]. This has the advantage of simplifying the neural architecture, as the number of layers can be significantly reduced. In such an architecture, one needs fewer neurons and can hence reduce the amount of computing effort required for training.

2. Method
Normally, image recognition takes place in the spatial domain. Using the Fourier-transform, data can be converted from the spatial domain to the frequency domain. There are discrete variants of the Fourier-transform, e.g. Discrete Cosine Transform (DCT), which we use in our experiments. A 2D image can be transformed into its DCT using the following equation:

$$X_{k_1,k_2} = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1,n_2} \cos \left[ \frac{\pi}{N_1} \left( n_1 + \frac{1}{2} \right) k_1 \right] \cos \left[ \frac{\pi}{N_2} \left( n_2 + \frac{1}{2} \right) k_2 \right]$$

(1)

where $k_1, k_2 = 0, \ldots, N_{1,2}$, and $N_1$ and $N_2$ are the numbers of pixels in horizontal and in vertical directions, respectively. In the 2D images of the DCT, the horizontal frequencies are presented from left to right and the vertical frequencies from top to bottom.

The transform is shown in the following Figures. The input consists of three images. They represent the class vertical bar, as shown in Figure 1. In the three images, the bars have an offset by one pixel. An exactly vertical bar has only horizontal frequencies. In the DCT representation, this is represented by pixels in the first row on top of the images.

![Figure 1. Vertical bars.](image)

![Figure 2. DCT of vertical bars calculated using equation 1.](image)

The DCT is based on the cosine function. In the transformed representation, the DCT values range between -1 and 1. In the DCT images, this is shown as follows: White for value 1 is, black for value -1, and gray for value 0, as one can see in Figure 2. The class Vertical bar is thus represented as frequencies in the top line of the DCT images. For classification, it does not matter whether the values are positive or negative. If a bar is moved, it can happen that at a point on the horizontal bar 1 has a negative value, for bar 2 a positive value appears. For bar 3 it is again negative. This is denoted by the white arrows in Figure 2. In the learning phase of neural networks, positive and negative values for one class at the same frequency disturb the learning. The weights of the neuron are strengthened by the positive and weakened by the
negative input. This has extinguishing effects in the learning phase of the NN. For this reason all values of the DCTs in our network are turned to absolute values, leading to the following equation:

\[ Y_{k_1,k_2} = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x_{n_1,n_2} \left| \cos \left( \frac{\pi}{N_1} \left( n_1 + \frac{1}{2} \right) \right) \cos \left( \frac{\pi}{N_2} \left( n_2 + \frac{1}{2} \right) \right) \right| . \]  

(2)

**Figure 3.** Absolute value DCT of vertical bars calculated using equation 2.

The images shown in Figure 3 are absolute value DCTs of the vertical bars, where value 1 is displayed as white and value 0 as black. Since vertical bars result in horizontal frequencies, the activities lie on the top row of the image, as is shown in Figure 3. In contrast, horizontal bars result in vertical frequencies. Hence the activities stand on the left side of the image. For diagonal bars, the result frequencies lie in diagonal direction. Because circles consist of evenly distributed frequencies in all directions, their spectral representations are co-centrically distributed frequencies around the upper left corner of the image. In traffic signs, the symbols are frequently made of the combinations of bars and circles, as shown in Figure 4.

**Figure 4.** Transformation results by absolute value DCT. Top row: Images of vertical bar, horizontal bar, diagonal bar and circle in spatial domain; Bottom row: The corresponding transformed images in frequency domain.

### 3. Experiments

A data set called German Traffic Sign Recognition Benchmark (GTSRB) has been used in our experiment. The size of the dataset is more than 50,000 images. It is divided into a training dataset and a test dataset. We use the online test dataset to calculate the quality of the NN. The images of the GTSRB were produced from videos of driving vehicles. The pictures show a total of 43 different traffic signs from Germany as you can see in Figure 5, i.e. they are assigned to 43 different classes.

**Figure 5.** Example images from GTSRB, with all 43 Classes.
The images of the GTSRB data set have an image size between 15×15 and 250×250 pixels. The images are not necessarily squared. They are livelike and have a very high variety in brightness and contrast. Since we do not normalize the size of the input signal within our neuronal network, we normalize the images ourselves to 40×40 pixel as part of a preprocessing.

For dark images, the brightness is increased, and for low-contrast images, the histogram is used to increase the contrast. After normalizing all the images, we name the processed data set as GTSRB Spatial Domain data set. It represents the images in space domain, which is split into a training and a test set accordingly. In a further step, we convert all images of the data set into frequency space using the absolute value DCT transform defined by equation 2. The input images are in color format with three channels.

We perform the DCT transform for each channel separately, and store the values again in a 40×40 pixel image in RGB color format. The transformed data set is named as GTSRB Frequency Domain. It represents the images in the frequency space.

We have performed tests with a neural network whose input layer has 40×40×3 = 4800 neurons, one hidden layer with a variable number of neurons, and an output layer consisting of 43 neurons, i.e. equals to the number of classes. For the experiments we vary the number of neurons in the hidden layer between 10 and 100 neurons. The neuronal net is trained with the respective training data set over 50 epochs. Proportional to the number of neurons in the hidden layer, the training times range between 5 and 30 minutes on a standard laptop. After training, the quality of the NN is evaluated using the test data set, based on the achieved classification accuracy. An accuracy value of 1 means that all images of the test data set have been correctly recognized by the neural network. For the purpose of comparison, tests have been performed using data set both in the GTSRB spatial domain and the GTSRB frequency domain. The achieved results are shown in Figure 6.

![Figure 6](image_url)

**Figure 6.** Accuracy achieved with different number of neurons in the hidden layers.

In the frequency domain we achieve an accuracy of 0.996, which means that 99.6 percent of the images in the test set are classified correctly. The recognition rate for the GTSRB frequency domain data set is always higher than the recognition rate for the GTSRB spatial domain data set. Especially with a small number of neurons in the hidden layer, the difference is significant. With only 40 neurons in the hidden layer we still achieve an accuracy of 95 percent in the frequency domain. If the number of neurons in the hidden layer is 10, it is not possible to recognize most classes in the GTSRB spatial domain, so that its accuracy is not measurable. In the frequency domain, the classification still works with 10 neurons.

In order to explain the functionality of the NN in the frequency space, we have carried out further experiments. Here, we have created synthesized representations of the traffic signs. They have the same dimension of 40×40 pixels, making the DCT of synthesized images comparable to that of real images. Traffic signs consist of a combination of symbols, usually a border such as a circle or triangle, and within it a number or a pictogram. For demonstration purpose, we have selected 10 different traffic signs, which are divided into three groups: Speed limits in the form of a red circle, danger signs in the form of a triangle and mandatory signs in the form of a blue circle with white arrows. The synthesized versions were decomposed into several
components, each as a partial symbol, as is shown in Figures 7, 8, 9. For example, the speed limit 70 (synthesized version) can be decomposed as a circle, the digit 7 and the digit 0.

In order to explain the frequency patterns in the GTSRB frequency domain dataset, we convert all symbols, pictograms and digits as well as the traffic signs combined from them into the frequency domain, using equation 2. Details on this can be found in Figures 7, 8, and 9.

Traffic signs for speed limits consist of the combination of a red circle on a white surface. Black digits on the white area show the maximum speed allowed in km/h. A red circle has edges in equal proportions in all directions. Therefore it is represented in the frequency domain by frequencies in all directions. The frequencies in the symbol image are all in the red excerpt of the image. The images from the GTSRB do not have a pure red excerpt due to many effects such as bleaching, lighting, camera calibration, but nevertheless the frequencies in the red excerpt are strongest. The frequency patterns of the circle can be found in the DCT from the GTSRB data set.

![Image of speed limit 70]

**Figure 7.** Speed limit 70, synthetic symbol and one real example from GTSRB, top row: spatial domain, bottom row: frequency domain.

The number 7 consists of a diagonal black bar, which gives a diagonal frequency, and in the upper part of the number a small horizontal bar, which gives a frequency in vertical direction. The digit 0 consists of a black circle with frequencies in all directions. In Figure 7 the frequencies of circle and digit 7 and digit 0 are denoted by white arrows. The recognition of numbers with neural networks in the frequency domain is already described in our previous research [12]. In the speed limit traffic sign 70 which consists of a combination of red circle and digit 7 and digit 0, the frequency patterns can be found again. The pattern is also present in the GTSRB image, which is shown in the last row of Figure 7. The NN learns these patterns.

Traffic signs for dangerous situations consist of a red equilateral triangle with a black pictogram on a white background. The red triangle forms a pattern in the frequency domain, by which all danger signs can be recognized. The black pictograms of the dangers for e.g. attention, skid danger, snow signs are different. The pictogram also have their own typical frequency pattern in the frequency domain. When the NN is trained on the DCTs of the GTSRB images, the neurons are trained on the combination of the frequency patterns of triangle and pictogram.

![Image of danger signs]

**Figure 8.** Danger signs, synthetic symbols and one real example from GTSRB, top row: spatial domain, bottom row: frequency domain.

Mandatory traffic signs consist of a blue circle with white arrow symbols on a blue background. The white symbols on a blue background form a frequency pattern in the complementary colour
yellow, i.e. in RGB colour format in the red and green extracts. The white arrow symbols also have their own characteristic frequency pattern in the DCT, which is recognized by NN.

**Figure 9.** Mandatory signs, synthetic symbols and one real example from GTSRB, top row: spatial domain, bottom row: frequency domain.

### 4. Conclusion and future work

This work deals with traffic sign recognition in the frequency domain. Based on the GTSRB test data set, a recognition accuracy of 99.6% has been achieved. We have also demonstrate that the recognition of traffic signs works better in the frequency domain than in the spatial domain. Particularly in cases of small models with limited number of hidden neurons, the model trained with features in frequency domain outperforms significantly its counterpart in spatial domain. In comparison to the spatial domain, for the same accuracy, it is possible to reduce the number of neurons if one works in the frequency domain, leading hence to the reduction of the overall computing power. Through the analysis of the frequencies of synthetic symbols, it is possible to explain why our classification approach works, making the neural methodology explainable in the frequency domain. In the future, we are going to verify our approach onboard a ground vehicle in the laboratory.

### 5. References

[1] Timofte, R., Zimmermann K., Van Gool, L.: Multi-view traffic sign detection, recognition, and 3D localisation, Machine Vision and Applications archive, Volume 25, pp. 633-647, 2014

[2] Mogelmose, A., Trivedi, M., Moeslund, T.: Vision-Based Traffic Sign Detection and Analysis for Intelligent Driver Assistance Systems, IEEE, Intelligent Transport. Sys., pp. 1484-1497, 2012

[3] Laguna, R., Barrientos, R., Felipe Blázquez, L., Miguel, L.J.: Traffic sign recognition application based on image processing techniques, IFAC Proceedings, Volume 47, pp. 104-109, 2014

[4] Larsson, F., Felsberg, M., P.-E. Forssen: Correlating Fourier descriptors of local patches for road sign recognition, IET Computer Vision Volume 5, pp. 244-254, 2011

[5] Stallkamp, J., Schlipfing, M., Sahmen, J., Igel, C.: Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition, Neural Networks, 2012

[6] Zakhloua, F., Stanciulescu, B., Hamdoun, O.: Traffic Sign Classification using K-d trees and Random Forests, Proceedings of Int. Joint Conf. on Neural Networks, San Jose, USA, 2011

[7] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-Based Learning Applied to Document Recognition Proceedings of the IEEE 86(11), pp. 2278-2324, 1998

[8] Sermanet, P., LeCun, Y.: Traffic Sign Recognition with Multi-Scale Convolutional Networks, 2011

[9] Qian, R., Yue, Y., Coenen, F., Zhang, B.: Traffic sign recognition with convolutional neural network based on max pooling positions, 12th ICNC-FSKD, 2016

[10] Shustanov A., Yakimova, P.: CNN Design for Real-Time Traffic Sign Recognition, Procedia Engineering, Volume 201, pp 718-725, 2017

[11] Bruna, J., Mallat, S.: Invariant Scattering Convolution Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 35, pp. 1872-1886, 2013

[12] Franzen, F., Yuan, C.: Visualizing Image Classification in Fourier Domain, ESANN 2019. Bruges, Belgium, 24-26 April 2019