IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 29, 2018, accepted November 24, 2018, date of publication December 4, 2018,
date of current version December 31, 2018.

Digital Object Identifier 10.1109/ACCESS.2018.2884826

Classification of Traffic Signs: The European Dataset

CITLALLI GÁMEZ SERNA AND YASSINE RUICHEK
Le2i FRE2005, CNRS, Arts et Métiers, University Bourgogne Franche-Comté, University of Technology of Belfort-Montbéliard, F-90010 Belfort, France
Corresponding author: Citlalli Gámez Serna (citlalli.gamez-serna@utbm.fr)

This work was supported by the Mexican National Council of Science and Technology (CONACYT).

ABSTRACT Classifying traffic signs is an indispensable task for autonomous driving systems. Depending on the country, traffic signs possess a wide variability in their visual appearance making it harder for classification systems to succeed. Either the classifier should be fine-tuned or a bigger collection of images should be used. In this paper, we introduce a real-world European dataset for traffic sign classification. The dataset is composed of traffic sings from six European countries: Belgium, Croatia, France, Germany, The Netherlands, and Sweden. It gathers publically available datasets and complements French traffic signs with images acquired in Belfort with the equipped university autonomous vehicle. It is composed of more than 80 000 images divided in 164 classes that at the same time belong to four main categories following the Vienna Convention of Road Signs. We analyzed the intra variability of classes and compared the classification performance of five convolutional neural network architectures.

INDEX TERMS Convolutional neural networks, dataset, traffic signs, traffic sign classification.

I. INTRODUCTION

Nowadays, Intelligent Autonomous Vehicles together with Advanced Driver Assistance Systems (ADAS) deal with the problem of traffic sign recognition. It is a challenging real-world computer vision problem due to the different and complex scenarios they are placed into. Some of the hard conditions include: illumination changes, occlusions, perspectives, weather conditions, aging and human artifacts to name a few. Therefore and because of the high industrial demand for autonomous vehicles, many studies have been published together with datasets from all over the world [1]–[10]. However, the systems are limited to the country and/or certain types of signs (shape, category).

Traffic signs provide crucial visual information in order to understand the proper driving conditions [11]. For example, they inform about speed limits, drivable lanes, obstacles, temporary situations, roadway access, restrictive areas, etc. Reasons why they are designed to be easily detectable, recognizable and interpretable by humans [6]. Standard shapes, colors, pictographs and text are used to define a meaning.

Nevertheless and besides the efforts to standardize traffic signs [12], there exists inter and intra variability between countries and between classes for specific traffic signs. For example, the inter variability is mostly seen between countries that do not follow a common convention [9] while intra variability is perceptible among places which agreed to follow one. In Europe, the Convention on Road Signs and Signals [12] established the common sizes, shapes and colors to be used but allows each country to choose its own symbols and inscriptions. Fig. 2 illustrates some examples of intra class variability where it can be seen that symbols do not only vary between countries but also inside each of them. Regarding the last issue, Croatia and France (Fig. 2 second row) use 2 symbols for pedestrian crossing sign in Danger category while Belgium has speed limit signs with and without adding the Km inscription. Germany also uses 2 different symbols in the pass-right class which belongs to Mandatory category (Fig. 2 fourth row). At the same time, the background color in some categories can vary as defined in [12]. For example, Croatia uses the two possible colors (yellow and white) for danger and prohibitory signs (Fig. 2, first and third rows) while the other countries stick to only one.

As mentioned earlier and due to the importance of traffic sign recognition, the research in this field has been popular and several methods that use selected hand-coded features as well as the ones which extract the features automatically have been proposed [13]. Among them, the most effective ones relying on CNNs architectures [6]. However, being able to recognize the same traffic sign in different countries is still a problem that in our knowledge, not many studies have

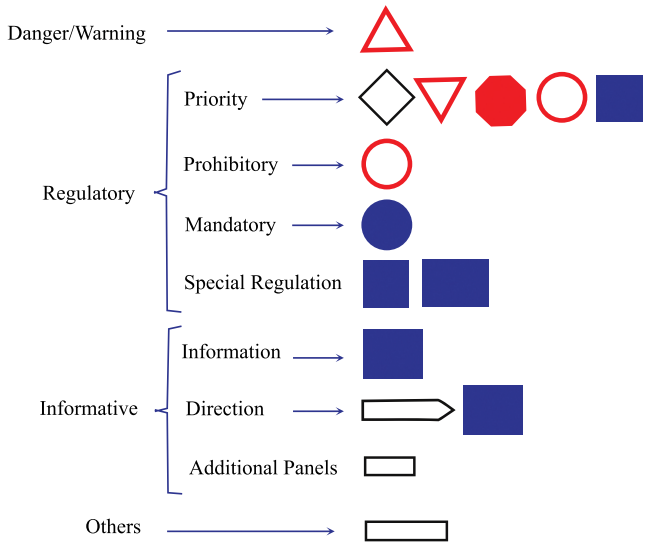

FIGURE 1. European traffic sign categories definition. From left to right: main category, subcategories and most common shapes.

FIGURE 2. Intraclass variability examples of European traffic signs.

addressed, specially in a continent (Europe) where countries are a few hours apart. In this paper we summarize our contributions to the following:

- A standard definition of 164 European traffic signs classes based on the Vienna Road Traffic Sign Convention [12]. The classes belong to 4 main categories and subcategories as seen in Fig. 1.
- A European traffic sign dataset that deals with intra class variability. It is composed of 82,476 images from 6 countries (Belgium, Croatia, France, Germany, Netherlands and Sweden).
- A comparative study of 5 CNNs architectures trained with our proposed European dataset and the German Traffic Sign Recognition Benchmark (GTSRB) [6].

The European dataset is available upon request and the pre-trained models can be found in `https://github.com/citlag/European-Traffic-Sings`.

The paper is organized as follows: Section 2 presents related work for traffic sign datasets and classification. Section 3 provides details about the dataset construction and class definitions. Section 4 describes the neural networks architectures used for training the datasets to continue with the analysis results in Section 5. Conclusions and future work are presented in Section 6.

II. RELATED WORK

The first work on traffic sign recognition was carried out in Japan in 1984 [14] and since then a broad number of works have been proposed to solve the problem through different techniques [15]. The most common ones are based on Support Vector Machines (SVM) [16], [17], template matching [18]–[20] and recently CNNs.

CNNs surpassed human performance on traffic sign classification [6], [15]. However, their architectures differ significantly from each other.

Even though traffic sign classification has been studied for decades, research works couldn't be compared until the German Traffic Sign Recognition Benchmark (GTSRB) [6] and the German Traffic Sign Detection Benchmark (GTSDB) [21] were proposed. Previously all research solutions have based their results on different public available datasets or on information acquired by their own.

The work of Abedin *et al.* [22] is an example of the formerly mentioned. They proposed the whole pipeline for detection and recognition. The recognition is carried out using SURF descriptors trained by an artificial neural network (ANN) with signs collected by themselves through video sequences in Bangladesh.

Islam and Raj [23] performed classification on 10 Malaysian signs through an artificial neural network with a 2-layer feed-forward and a softmax classifier. Each class was composed of 100 samples dividing them into 70%-15%-15% for train, test and validation sets respectively. The signs were captured on roads and highways during different daytimes and weather conditions. Their dataset was also used by Lau *et al.* [8] to compare 2 classification methods, a Radial Basis Function Neural Network (RBFNN) and a CNN.

Li *et al.* [24] proposed a convolutional neural network (CNN) to detect and classify U.S speed limit signs [9]. Their network is based on a modified version of R-CNN [25] for the detection and a Cuda-convnet [26] for the classification. They claim to achieve 93.89% mean AUC [24] for 4 classes (No Turn, Speed Limit, Stop and Warning).

In the same manner, Jung *et al.* [10], collected and classified 6 types of traffic sings in South Korea. The training procedure was performed with LeNet-5 CNN architecture [27] predicting correctly 16 traffic signs on the road within an observable range.

Yang *et al.* [16] went beyond all the previously mentioned works, classifying not only the respective traffic sign classes but also their superclass (categories). Their system is based on 4-class SVM classifiers with RBF kernel using Color HOG features to detect the traffic sign categories. Then, three CNNs are used to perform real time traffic sign recognition. Each CNN contains two convolutional layers followed by

sub-sampling layers, plus a fully-connected MLP in the last two layers. Their method was trained and evaluated with the GTSDB [6].

Aghdam *et al.* [15] designed a CNN architecture inspired by Cireşan *et al.* [28] and compared their work to 3 other networks [28]–[30] reducing by 65%, 63% and 54% the number of training parameters respectively. Their proposed CNN is trained with the GTSRB [6] and fine-tuned with the Belgium dataset [4], [7] in order to prove that their architecture is not only efficient, but also transferable.

Recent work proposed by Li and Wang [31] manages traffic sign detection and recognition. Their traffic sign recognition CNN uses different asymmetric kernels in order to reduce the number of convolutional operations. Additionally, they fused different spatial information using an inception module by concatenating the output of 2 CNN branches along the channel axis. The CNN model was trained with the GTSRB proving to be effective and robust obtaining 99.66% accuracy.

In our study, we will train different CNNs architectures on the same datasets applying a common data preprocessing step and number of epochs in order to provide a fair comparison of traffic sign recognition approaches.

III. DATASET

The images of our proposed European dataset are composed of public available datasets and of sequences recorded in Belfort, France and surroundings during Spring and Summer from 2014, 2015 and 2018. The sequences are composed of urban and rural environments and cover daytime and sunset conditions. The public datasets are composed of different scenarios (urban, rural, highway) mostly captured during daytime.

A. DATA DEFINITION

In order to standardize traffic signs, a lot of efforts have been made since 1909 to establish a common structure. Yet, in 1968 in Vienna, the United Nations Economic Commission for Europe (UNECE) established the Convention on Road Signs and Signals which entered into force on 1978 [12]. Currently, 62 countries follow it with small variations in colors, pictographs and text.

The proposed dataset is divided into 4 relevant categories and subcategories (see Fig. 1) following the Vienna Convention on Road Signs and Signals [12]:

1) Danger warning signs. Warn road-users of a danger on the road and inform them of its' nature.
2) Regulatory signs. Inform road-users of special obligations, restrictions or prohibitions with which they must comply.
3) Informative signs. Guide road-users while they are traveling or provide them with other information which may be useful.
4) Others. Added class to inform road-users about important situations.

Regulatory and informative categories have subcategories that other datasets have considered for defining their traffic signs classes [6], [22]. For example, the German Traffic Sign Recognition Benchmark (GTSRB) divided the traffic signs in 3 categories: 1) Danger, 2) Prohibitory, 3) Mandatory and include Other as an irrelevant category. Categories 2 and 3 belong to the Regulatory category defined in the Vienna Convention.

FIGURE 3. Examples of French directional signs.

As we will explain in the following subsection, most of the public available datasets include only Danger or Mandatory signs. In the contrary, our proposed dataset considers a complete definition (including Informative signs) for traffic sign classification. The reason behind this, resides on facilitating the tedious task to recognize traffic signs which are not only composed of standard shapes with pictographs, but also to be able to recognize more complex signs for later interpretation. Directional signs are an example of that. They are composed with text and shapes to indicate certain direction (Fig. 3).

At the same time, additional panels provide complementary information to interpret traffic signs correctly. The broad selection of traffic signs resulted in 164 classes shown in Fig. 4.

A CSV file describing class names together with their categories and subcategories they belong to, is provided upon request together with the respective European dataset. In the following subsections (III-B and III-C), we provide more information about how our proposed dataset was built.

B. DATA COLLECTION - UTBM DATASET

We use the equipped vehicle of the UTBM laboratory to capture several sequences of images around university campus. The vehicle is equipped with multiple sensors: a Bumblebee 3D stereo vision camera mounted on the top, a Real Time Kinematic (RTK) GPS sensor and several Light Detection and Ranging (LIDAR) devices. Data collection with the Bumblebee camera was performed with automatic exposure control and a frame rate of 16 fps. The sequences are captured during daytime covering urban environments with a resolution of 1280×960 pixels. Images for annotation were chosen every 3.5 meters.

In addition to the previous sequences, we captured images in the surroundings of Belfort with a conventional Canon Eos M camera with automatic exposure and focusing point. Driving scenarios cover urban and rural environments. The resolution of the video is 1920×1080 pixels at a frame rate of 24 fps. Images for annotation were chosen every 6 fps.

FIGURE 4. Random representation of the 164 classes in the European traffic sign dataset. (a) Danger warning signs. (b) Regulatory - Priori. (c) Regulatory - Prohibitor. (d) Regulatory - Mandatory. (e) Regularory - Special regulation. (f) Informative - Information. (g) Informative - Direction. (h) Informative - Additional panels. (i) Others.

Data annotation was performed using the Training Image Labeler tool from MATLAB. Regions of interest of all seen traffic signs are labeled manually, cropped and saved in PPM file format.

Every labeled sequence was recorded on a single tour in different days and places. Some sequences were recorded several times at different seasons and lighting conditions, this with the aim to include variance in the image dataset. All the recordings belong to France subset dataset.

Details about the number of images and traffic signs labeled will be provided in the next subsection, together with the other datasets.

C. DATA ORGANIZATION

The European traffic sign dataset (ETSD) gathers public available datasets from 6 countries: Belgium with the KUL Belgium Traffic Signs dataset [7], Croatia with MASTIF datasets [2], France with the Stereopolis dataset [5], [32], Germany with the well-known German TSR Benchmark (GTSRB) [6], Netherlands with the RUG Traffic Sign Image Database [1] and Sweden with the Swedish Traffic Signs Dataset (STS dataset) [3].

In order to complement the already provided classes and add the missing ones (signs from Informative and Others categories), we labeled the German Traffic Signs Detection Benchmark (GTSDB) [21] and the RUG [1] datasets.

A detailed description of the traffic signs per dataset is described below:

1) Belgium. KUL Belgium Traffic Signs dataset [7] is a dataset for classification where each image represents a sign with 10% offset. They are cropped according to the ground truth information to obtain only the Regions of Interests (RoIs). The original dataset is divided into 62 classes with 4,561 images for training and 2,528 images for testing. For the ETSD, its class 35 was divided into 3 classes (3 Mandatory signs), class 32 was divided into 6 classes (6 Prohibitory signs) ending up with 69 European classes.
2) Croatia. MASTIF datasets [2]. There are 3 datasets collected in different years (2009-2011). Each sign was annotated 4-5 times at different distances from the car. In order to distinguish images from each dataset, we put as prefix the dataset year for each image name.
    a) 2009. It was the richest classification dataset containing 6,423 signs already representing the RoIs. This dataset is composed of 97 classes from which we divided "b31" class into 7 prohibitory classes, merged 19 classes into 6 and dropped 2, resulting in a total of 88 classes with 6,411 signs for our dataset.
    b) 2010. It is a detection dataset composed of 3,862 images of resolution 720×576 pixels. Signs were cropped according to the ground truth obtaining 5,184 RoIs divided into 88 classes. We separated "A17" class into 2 dangerous classes, "b31" into 5 prohibitory classes, merged 18 classes into 4 and dropped 1 class. The total number of classes resulted in 68 in our dataset.
    c) 2011 dataset provides 1,013 images with the same resolution as in 2010 dataset. 1,429 traffic signs were cropped and distributed into 53 classes according to the ground truth. We divided "b31" class into 2 prohibitory classes, "c35" into 2 Special Regulation classes, "e19" into 2 Additional Information classes and merged 12 classes into 2. A total of 41 classes emerged for our European dataset.
3) France.
    a) Stereopolis dataset [5], [32] includes images acquired in Paris, France, grabbing a picture every 5 meters. It is made of 847 images of resolution 960×1080 pixels. We cropped the traffic signs with the ground truth provided and obtained 271 road signs divided into 10 classes. We split "a13a" class into 2 Danger classes and "b6a1" into 2 Prohibitory classes resulting in 12 classes for our dataset.
    b) UTBM dataset is composed of 86 classes with 2,631 signs. Image resolution varies as described in Section III-B. A total of 1,863 images were labeled comprising 4 sequences captured with the Bumblebee 3D camera and 6 sequences with the Canon Eos M.
4) Germany.
    a) GTSRB [6] dataset contains a total of 51,839 images comprising 43 classes. The training and testing sets were used for our European dataset as their original definitions.
    b) GTSDB [21] dataset comprises 900 full images with a resolution of 1360×1024 pixels. Originally 43 classes are labeled according to GTSRB, but we extended the dataset labeling manually with RoIs the sings not considered in the image. We obtained 1,187 RoIs composed of 46 additional classes. In total, we obtained 89 classes for our dataset (46 labeled classes + 43 original classes). We named these images with the prefix "GTSDB_" followed by a unique index number.
5) Netherlands. RUG Dataset [1] contains 48 images of size 360×270 pixels in PNG format. Originally, the dataset was used to classify 3 classes (pedestrian crossing, compulsory for bikes and intersection). We labeled all traffic signs seen including the 3 classes as no ground truth is provided. 75 signs belonging to 12 classes were cropped and saved with the prefix "neth_" followed by a unique number.
6) Sweden. STS Dataset [3] provides 2 sets (Set1 and Set2) with a total of 3,777 images labeled manually with RoIs. 6,363 signs were extracted from the ground truth full images divided into 19 classes. We separated

FIGURE 5. Examples of some classes that contain intra class variability. Letters in the images are the initial of the country name they belong to. For example, B stands for Belgium, C for Croatia, etc.

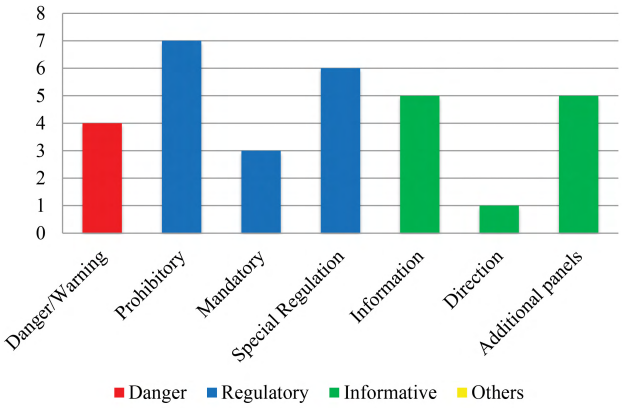

FIGURE 6. Number of classes per category with less than 10 signs.

TABLE 1. Relation of classes and total number of traffic sings by country.

Country	Dataset	Original classes	European classes	Training images	Testing images	Total images
Belgium	KUL	62	69	4561	2528	7089
Croatia	MASTIF-2009	97	88	4568	1843	6411
	MASTIF-2010	88	68	3694	1490	5184
	MASTIF-2011	53	41	1037	389	1426
France	Stereopolis	10	12	203	68	271
	UTBM	86	86	1878	753	2631
Germany	GTSRB	43	43	39209	12630	51839
	GTSDB	43	89	859	328	1187
Netherlands	RUG	3	12	58	17	75
Sweden	STS-Set1	19	52	2092	829	2921
	STS-Set2	19	52	2409	1033	3442

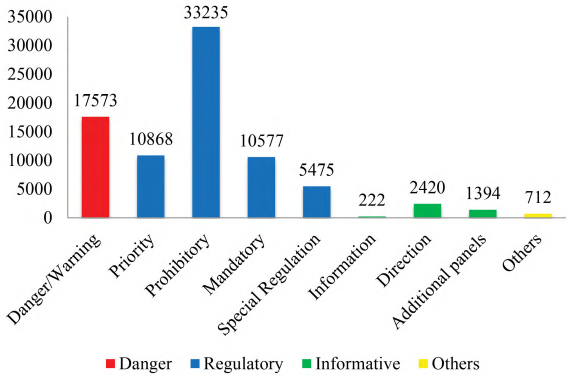

FIGURE 7. Number of images grouped by category and sub-category of the proposed European dataset.

class "OTHER" into 34 classes. We named the new signs with the prefix "set1_" and "set2_" followed by a unique index for each corresponding set.

The naming convention used for each sign is as follows: *X_originalName.ppm* Where X is replaced with a "B" for Belgium, "C" for Croatia, "F" for France, "G" for Germany, "N" for Netherlands and "S" for Sweden. We provide a CSV ground truth file to reference the given names with the original files in the necessary cases.

Traffic sign sizes vary between 6 and 780 pixels w.r.t the longer edge. Bounding boxes are not necessarily squared due to the nature of the traffic signs. Our final dataset takes care of intra class variability, as can be seen in Fig. 5, comprising 82,476 signs with 164 classes. We split the dataset into training and testing considering 70 - 30 % ratio only in the cases where the class contains at least 10 signs. If the class contains less than 10, we consider 100% for training. Fig. 6 illustrates the distribution by category of the classes with less than 10 signs. Moreover, we carefully select random images per country for testing. In this way, we make sure the testing phase validates the accuracy for European signs and not for a specific country. The training and testing sets are ordered by class. Table. 1 shows a summary of the classes and images considered for training and testing per country.

It is important to mention that even though the dataset is unbalanced as seen in Fig. 7 (with some classes containing less than 10 images), the purpose of its definition was to expand and establish a common notation for traffic signs classification based on the Vienna Convention [12]. The aforementioned will allow former researchers to contribute with more data in some categories or specific classes.

The European Traffic Sign dataset (ETSD) contains images that deal with occlusions, different lighting conditions, motion blur, human made artifacts and perspectives (see Fig. 8). Although, it does not contain signs captured during night or extreme weather conditions, data-augmentation is a well-known and applied technique in the literature for classification tasks to generate data and reduce the effect of over-fitting [15], [33]. Normally, the synthetic images are generated applying different transformations, simulating dark or bright scenarios, adding noise, blur, etc. to compensate the lack of information and avoid data collection and labeling. In this way, if data-augmentation is applied, we can say that our proposed dataset possesses the necessary characteristics as illustrated in Fig. 8 to consider it more robust for classification tasks. We will show in Section V its advantages along with applying data-augmentation.

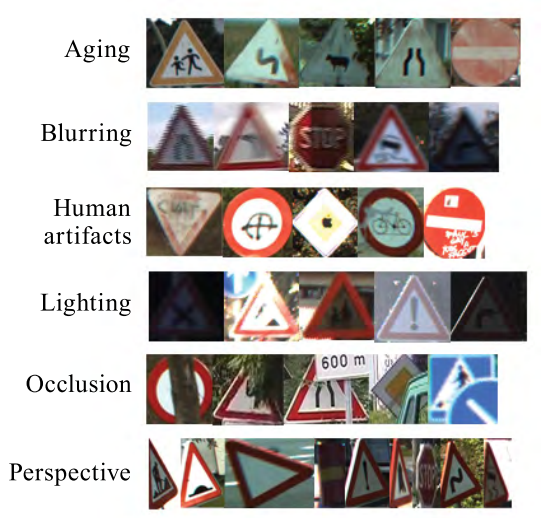

FIGURE 8. Examples of challenging signs.

IV. CONVOLUTIONAL NEURAL NETWORKS (CNNs) FOR TRAFFIC SIGN CLASSIFICATION

It has been proven that CNNs are capable to solve problems with really high accuracy compared to human performance [28], [34]. Since the German Traffic Sign Recognition Benchmark [6], a lot of works were proposed to deal with traffic sign classification through different machine learning methods [10], [15], [28], from which CNNs outperform the others. As traffic sign classification is in high demand for the automotive industry, a lot of efforts have been made to achieve real time classification [10], [28].

We will describe 5 CNNs that achieve the best performances in the state of the art regarding Traffic Sign Classification.

A. LENET-5

LeCun *et al.* [27] proposed the well-known LeNet-5 convolutional neural network that is mostly used for handwritten recognition. Besides it was introduced in 1998, it became popular to solve other problems due to its simple and efficient architecture. It is composed of 7 layers, 3 Convolutional layers followed by Sub-sampling layers (except in the last), 1 Fully connected layer and the final output layer composed of Euclidean RBF units. The input size for this network is 32×32 pixels.

Jung *et al.* [10] used LeNet-5 to classify 6 types of Korean traffic signs obtaining an accuracy of 100% recognizing correctly 16 signs while driving on the KAIST campus road. As the results were promising in their study, we also trained the network with our proposed dataset for comparison.

B. IDSIA MODEL

Cireşan *et al.* [28] based their work on combining several Deep convolutional Neural Networks (DNN) columns to form a Multi-column DNN (MCDNN). Their DNN network is composed of 2 Convolutional layers followed by Max-pooling layers. At the end, 2 fully connected hidden layers are used to pass the output to a final fully connected layer with 6 neurons to perform classification. They used a scaled hyperbolic tangent activation function for convolutional and fully connected layers. Their net takes as input 2 images of 48×48 pixels and performs some distortions in each column to average at the end the final predictions of each DNN.

Aghdam *et al.* [15] trained this model with the GTSRB dataset and obtained an accuracy of 98.52% performing data-augmentation for the training set.

C. URV MODEL

Aghdam *et al.* [15] made a comparative study between methods using the GTSRB [6]. Their CNN based on Cireşan *et al.*'s work [28] demonstrated in their results, that their network is able to reduce complexity and computational time, improving accuracy compared to the one of Ciseran et al. Their Network is based on 3 convolution-pooling layers and 2 fully connected layers with a dropout layer in between to avoid over-fitting. ReLU activations [35] after each convolutional layer and after the first fully-connected layer are applied. Their network takes as input a 48×48 RGB image and classifies it into one of the 43 traffic sign classes of the GTSRB dataset. They claim to achieve 98.94% accuracy performing data-augmentation for the dataset.

D. CNN WITH ASYMMETRIC KERNELS

Li and Wang [31] based their CNN design on convolutional layers using asymmetric kernel sizes to replace the usual symmetric $n\times n$ kernel (e.g. 3×3,5×5,7×7), with asymmetric ones defined by $n\times1$ and $1\times n$ in some convolutional layers. This replacement decreases the number of convolutional operations making the network more efficient. Their CNN architecture is composed of 3 convolutions with symmetric kernels, 6 convolutional layers with asymmetric ones (7×1, 1×7, 1×3, 3×1, 1×7 and 7×1), and 2 fully connected layers. Each of these layers except for the last one (Softmax classifier) are followed by Batch Normalization [36] and ReLu activations [35]. They used an inception module with asymmetric kernels after the third convolution to learn different spatial information. The last two convolutional layers use symmetric kernels. Dropout technique [37] is used by the authors to avoid over-fitting. As they trained the network with the GTSRB, the output is set to 43 and input size to $48\times48\times3$ (RGB image).

The performance of this CNN achieved 99.66% accuracy on the GTSRB test set trained with data-augmentation for 200 epochs. Despite the use of asymmetric kernels to decrease the complexity of the network, the number of parameters for this architecture is still high compared to the others studied here. A comparison of this metric will be provided in the next Section.

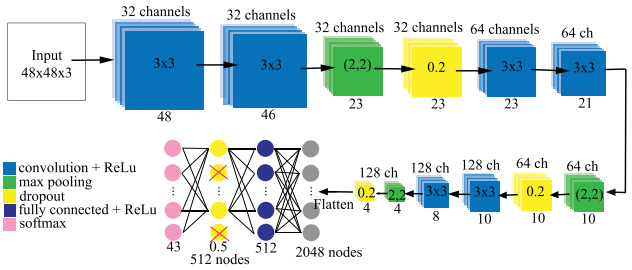

FIGURE 9. Architecture drawn from Chilamkurthy proposal [42].

E. CNN 8-LAYERS

Besides the previously mentioned architectures, we decided to train a classifier that does not represent a really deep network but competes with the high accuracy in the state of the art. Networks composed with deep architectures (several hidden layers) have proven to provide the best results (Inception [38], VGG16 [39], ResNet [40]), but their complexity kills the computational time. For this reason, simple networks are used considering information preprocessing and data-augmentation [41] if the dataset does not contain enough examples for the learning phase.

Chilamkurthy [42] worked on the traffic sign image classification problem. Even though he did not mention in which Network he based his work, we could see that his proposal is like a VGG architecture [39]. There are blocks of Convolutional layers, activated by ReLu function [35], followed by Max Pooling and additionally Dropout layers. His architecture can be seen in Fig. 9. Different from VGG architecture, he added dropout layers with a range of 0.2 after each block of convolutional layers and, in the same way, after the fully connected layer with a range of 0.5. Dropout layers are used to prevent over-fitting and to make the network learn robustly its parameters [37].

Chilamkurthy reported 97.92% and 98.29% accuracies without and with data-augmentation respectively using the GTSRB dataset. His network takes RGB images of size 48×48 pixels as input, while transforming them to HSV color space and performing histogram equalization in the V channel.

A comparison of all the previously mentioned architectures will be performed in the next Section using the GTSRB and our European dataset.

V. RESULTS

In order to perform dataset comparison between different networks, we trained the models described in Section IV with the GTSDB and our proposed European dataset. All models are trained in GPU mode using a NVIDIA GeForce GTX1080Ti with 11GB of memory, an IntelCore i7K-8700K (6 cores 12 threads, 12 Mb cache memory) processor and RAM of 32GB. The learning process varied according to the complexity of each model.

The URV model proposed by Aghdam *et al.* [15] and the IDSIA model proposed by Cireşan *et al.* [28] are implemented in the Caffe framework.[1] The input image of both models is an RGB image of 48×48 pixels. We trained both models with the original parameters as stated by Aghdam *et al.* [15] changing only the batch size from 100 to 128 and the number of iterations to make it learn for the equivalent of 40 epochs (11500 iterations for the GTSRB and 17500 for the European dataset). In the same way, the test iterations are modified according to the validation dataset: 31 for the GTSDB and 48 for the European one.

TABLE 2. Accuracy percentage results obtained on the GTSRB and European test sets. The input size refers to image "width×height×channels", while the number of parameters is presented in millions (M) and time in milliseconds (ms).

Model	Input size	GTSRB		European		Time
		Parameters	Accuracy	Parameters	Accuracy	
LeNet-5	32x32x1	0.13 M	89.1%	0.35 M	89.8%	0.0067 ms
IDSIA	48x48x3	1.54 M	94.62%	1.58 M	95.82%	0.6 ms
URV	48x48x3	1.12 M	96.1%	1.16 M	96.53%	0.61 ms
CNN asymmetricK	48x48x3	2.92 M	97.88%	2.95 M	98.48%	0.39 ms
CNN 8-layers	48x48x3	1.48 M	98.52%	1.51 M	97.88%	0.15 ms

Towards a fair comparison, we normalize the European dataset subtracting the mean image like it is done for the GTSDB dataset. The results presented in [15], based their accuracy on augmented-data carried out with 12 transformations (see paper [15] for more details).

UVR and IDSIA models were trained without performing any data-augmentation or preprocessing. This with the aim to evaluate the performance of the pure models. We use 10% of the training sets for validation (3921 images for GTSDB and 6055 for European dataset). The results obtained can be seen in Table. 2, where the classification accuracies presented come from evaluating the models on the test sets.

In the same manner, we trained the models: LeNet-5 [27], the CNN with asymmetric kernels proposed by Li and Wang [31] (CNN asymmetricK), and the model proposed by Chilamkurthy [42] (CNN 8-layers) implementing some changes to increase the model accuracy. All these architectures are implemented in the Tensorflow framework.

Due to the nature of LeNet-5 [27], images are converted into gray-scale and resized to 32x32 pixels. We set the batch size to 128 and the number of epochs to 40 to train on both datasets. The rest of the training parameters are left unchanged according to the original implementation.[2]

Since the outputs of LeNet-5 fully connected (FC) layers are reduced from 400 (total number of neurons after the pooling layer of the second convolution) to 120-84-N (output classes), we had to modify the number of neurons of the FC layers in order to be able to classify 164 classes for our proposed European dataset. For that, we changed the first output of the first FC to 300 and the second one to 200. In this way, we are able to classify 164 classes as the desired output. Results are available in the Table. 2.

[1] https://github.com/pcnn/traffic-sign-recognition
[2] https://github.com/sujaybabruwad/LeNet-in-Tensorflow

Furthermore the model of Li and Wang [31] was trained as well for 40 epochs on both datasets. Images are resized to 48×48 pixels keeping the 3 color channels (RGB). Different from the authors, we applied a preprocessing step converting the image to HSV color space and equalizing the V channel. Preprocessing methods are used to normalize the image and give better contrast [30]. The mean image was also subtracted in each dataset. We use the model parameters as proposed by the authors and the results obtained are shown in Table. 2. We refer to this model as CNN asymmetricK.

Moreover, and in order to improve the accuracy of 97.92% reported by Chilamkurthy [42] on the GTSRB, the CNN 8-layers model was modified adding 1) L2 regularization on each convolutional and fully connected layers and 2) Batch Normalization after each convolutional layer and before the ReLu activations. The effect of Batch Normalization [36] has proven to make the network learn robustly normalizing the input parameters in each batch at each layer to reduce their covariance shift. Regularization, in the other hand, helps reduce over-fitting adding a penalty in the loss function to combat high variance. L2 regularization was added with a value of 1e-4. The optimizer was changed from Stochastic gradient descent to Adam [43] with an initial learning rate of 1e-3 and a regularization coefficient of 1e-3. Batch size was set to 128 and the input to 48×48 pixels with 3 channels (RGB image). In the same way as with CNN asymmetricK model, the input image is preprocessed as mentioned earlier. The model was trained for 40 epochs validating the training process using the validation set (10% of the training set) which stops the training when over-fitting occurs (validation loss starts increasing and validation accuracy start decreasing). In this way, we make sure the CNN model learns correctly. For example, while training the CNN 8-layers model with the European dataset, it stops learning at epoch 16 obtaining a testing accuracy of 98.52% (see Table. 2).

The changes implemented in the CNN 8-layers model made the network improve its accuracy 0.6% from 97.92% reported by Chilamkurthy [42] to 98.52% on the GTSRB dataset without performing data-augmentation. Besides the accuracy improvement, and due to the addition of Batch Normalization layers, the number of parameters also increased from 1.36 to 1.48 Millions in reference to the GTSRB dataset. Nevertheless and regardless the increase of complexity, the modified CNN 8-layers model is still competitive in accuracy and time compared to the others reported in Table. 2.

The accuracies obtained on both datasets (Table. 2) are similar, varying from 0.43% - 1.2% between the GTSDB and European datasets, no matter the model used. Hence we can say that models are stable and results depend on the dataset itself.

The processing time for each model depends on its number of parameters and the framework used. For instance, the processing times presented in Table. 2 are computed to predict the traffic sign class of a single image in GPU mode. Essentially, we can see that the models implemented in the Caffe framework (IDISA and URV) are relatively slower than the ones implemented in Tensorflow. For example, the IDSIA model (Caffe) has 1.54 Million parameters and takes 0.6 milliseconds to make a prediction, while CNN asymmetricK model (Tensorflow) has 2.92 Million parameters and takes 0.39 milliseconds (around 40% faster).

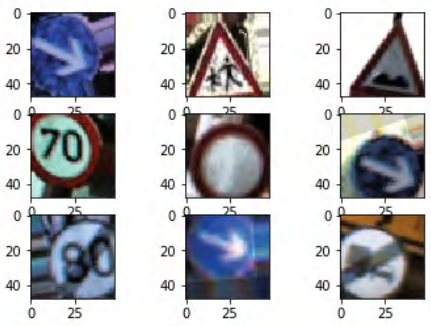

FIGURE 10. An example of some augmented traffic signs from the European dataset.

As mentioned before, techniques like data-augmentation also help to improve the accuracy of a classifier without acquiring and labeling more data. We applied this technique with the 2 models which obtained the best accuracies on the test datasets: the CNN asymmetricK model [31] and the modified CNN-8 layers model. Luckily, as these models are implemented in Keras using Tensorflow as back-end; Keras provides an option to perform real-time data-augmentation with its ImageDataGenerator class. We considered the following 5 transformations:

1) Width shift = +−4 pixels
2) Height shit = +−4 pixels
3) Scaling = [0.8,1.2]
4) Shear = [0,0.1] radians
5) Rotation = +−10 degrees

Besides that transformations, histogram equalization is also considered as data preprocessing. For this, the exact same procedure is applied as stated by Chilamkurthy [42]. Some examples of augmented images can be seen in Fig. 10.

Considering that the models were previously trained without data-augmentation, we used their pre-trained weights as initializers for the new training procedures. This technique is also called transfer learning and avoids learning everything from scratch. Normally, it is more common to use it with deep architectures which were trained on huge amount of data to adapt the model to a new output with less training examples [44]. In our case, we used it as initialization for the architectures to continue learning with the new generated data. The training parameters were left unchanged as defined previously and only the number of epochs was set to 50 for both CNN models. Table. 3 shows the accuracies obtained on the test sets on each datasets. These testing accuracies with the CNN asymmetricK model [31] are improved by 1.49% in the GTSDB dataset and 0.41% in the European dataset, while with the modified CNN 8-layers model, they are improved by 0.85% and 1.11% respectively.

TABLE 3. Accuracy percentage results obtained on the GTSRB and European test sets without and with performing data-augmentation.

Model	GTSRB		European	
	Original	Augmentation	Original	Augmentation
CNN asymmetricK	97.88%	99.37%	98.48%	98.89%
CNN 8-layers	98.52%	99.37%	97.88%	98.99%

The average human performance for detecting traffic signs on the GTSRB dataset is 98.84% as reported in [6]. Both CNNs trained in this study with data-augmentation surpassed the human performance with both datasets. For the URV model proposed by Aghdam *et al.* [15], we obtained 96.1% accuracy while they reported 98.94% applying 12 transformations as data-augmentation. With this in mind, we can affirm that a classifier learns more robustly if the dataset comprises a wide variety of data situations.

Due to the fact that the proposed European dataset comprises a wide range of situations and possesses a larger number of training data; most of the accuracy results for each model are improved comparing them to the accuracies obtained on the GTSRB dataset (see Table. 2). Nevertheless, even with data-augmentation, the models could not achieve more than 99% accuracy in the European dataset (Table. 3). We will analyze the predictions in both datasets with the 2 CNN models trained with data-augmentation to find out the reasons that made the classifiers failed.

Table. 4 presents the relations of the incorrect classes by category per model and dataset. We present the errors by category because it is the common way in the literature [6], [13], [16], [21] to compare the differences between groups of classes with similar characteristics (shape, color). For the European dataset (Table. 4), we can see that the CNN asymmetricK model has troubles in learning the categories Danger, Prohibitory, Special Regulation and Direction with more than 1% error, while the CNN 8-layers model has troubles with Danger, Special Regulation, Information, Direction and Additional panels with more than 1% error as well.

A deeper analysis for the incorrect predictions on the European dataset is performed to find out the characteristics of the traffic signs that make the classifiers failed.

The first intuition for bad predictions was image size and aspect ratio. We counted the number of misclassified signs that were 1) rectangular or squared and 2) big or small. A squared sign is considered if its aspect ratio falls between 0.9 and 1.1, while small signs are considered if the image has less than 255 pixels. The latest parameter is set taking into account that the smallest image size on the GTSRB is 15×15 pixels (255 pixels).

As a reference, the total number of incorrect predictions with the CNN asymmetricK model is 244, while for the CNN 8-layers model is 222. Fig. 11 shows the relation of the incorrect predictions according to the parameters previously mentioned. There, we can see that only a few signs are small in each of the classifiers (8.2% for CNN asymmetricK model and 13.06% for the CNN 8-layers model), while almost half of the incorrect predicted signs are rectangular with both classifiers. We considered these metrics because: 1) when the image size is small, even for humans, it is hard to distinguish the correct class; and 2) when the image is rectangular, the classifier resizes it to a squared size suffering information loss.

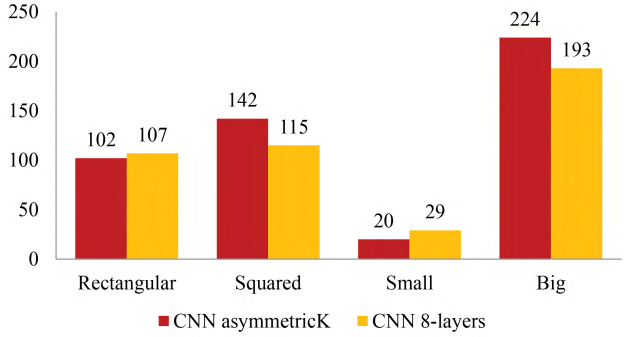

FIGURE 11. Image size analysis for the incorrect predictions on the European test set. Results are obtained with the CNN models trained with data-augmentation.

In the same way, we analyzed the predicted probabilities for the misclassified signs. We consider as uncertain predictions the ones which are incorrect and predicted with a probability equal or bigger than 0.9. The most uncertain predictions obtained with the CNN asymmetricK model were $107/244 = 43.85\%$ while for CNN 8-layers model were $65/222 = 29.28\%$. This kind of analysis can help a classifier refuse the prediction if the confidence probability is less than a certain threshold, however in this approach is not applicable. As we are interested in the visual characteristics that make the signs difficult to classify correctly, we inspected all the incorrect predictions. Fig. 12 and Fig. 13 illustrate some of the incorrect predictions for the CNN asymmetricK model [31] and CNN 8-layers model [42] respectively.

After the visual inspection, we found that most of the misclassified signs possess the following characteristics:

- Strong motion blur.
- Incomplete signs (cropped).
- Occlusions.
- Strong shadows or highlights.
- Strong perspectives.
- Human added artifacts.
- Poor image quality.
- Aging.
- Very different aspect ratios (rectangular signs).

Most of the errors in Danger and Regulatory categories are due to the characteristics listed above. For the Informative category, the misinterpreted signs are mostly due to their visual complexity and to the very different aspect ratios. Informative signs contain text, which by nature, makes them the hardest ones to recognize. At the same time, their very

TABLE 4. Error percentage predictions by category on the GTSRB and European test sets. Results are computed from the CNN models trained with data-augmentation.

Category	GTSRB			European		
	Signs in GT	CNN asymmetricK	CNN 8-layers	Signs in GT	CNN asymmetricK	CNN 8-layers
Danger	2790	1.25%	0.39%	4626	1.75%	1.36%
Priority	1680	0.00%	0.83%	2946	0.17%	0.20%
Prohibitory	6390	0.66%	0.83%	8625	1.18%	0.94%
Mandatory	1770	0.17%	0.11%	2818	0.43%	0.64%
Special Regulation	-	-	-	1550	1.35%	1.35%
Information	-	-	-	59	0.00%	3.39%
Direction	-	-	-	706	2.55%	3.12%
Additional panels	-	-	-	392	0.77%	1.79%
Others	-	-	-	208	0.96%	0.96%
Total	12630	0.63%	0.63%	21930	1.11%	1.01%

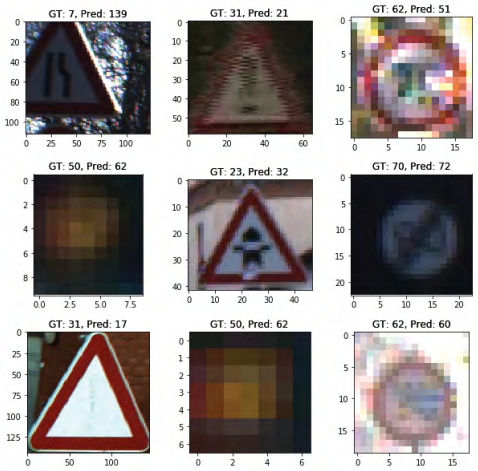

FIGURE 12. Random sample of incorrect predictions of the European dataset with the CNN asymmetric model trained with data-augmentation.

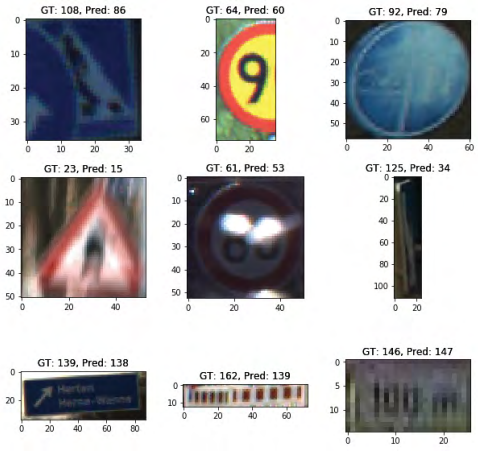

FIGURE 13. Random sample of incorrect predictions of the European dataset with the CNN 8-layers model trained with data-augmentation.

different aspect ratios conduce them to information loss once the classifier resizes them to a common input shape, normally, a squared shape. For example, in the Direction sub-category, most of the errors reside on confusion of class 139 (Direction to place) with class 138 (Advance directional signs) and vice-versa (see Fig. 13) due to the fact that both contain text and their appearances vary a lot.

Interestingly, no matter how many conditions our proposed European dataset considers (Fig. 8), there will always be hard situations for the classifiers to learn. In order to overcome this issue, image processing techniques can be used to enhance the visibility of an image and data-augmentation can be applied to improve the learning process generating more samples with different transformations.

In summary, our proposed European traffic sign dataset proved to be more robust than the GTSRB dataset with the 5 CNN architectures trained on (Table. 2) making it reliable and more complete for traffic sign recognition.

VI. CONCLUSION

We proposed a traffic sign European dataset which deals with intra-class variability from 6 countries (Belgium, Croatia, France, Germany, Netherlands and Sweden). Such characteristic is a crucial aspect for autonomous vehicles when driving from one country to another, since a classifier does not perform properly when traffic signs (pictographs or text) are slightly different from each other [15]. In Europe, this is a vital issue considering that countries are relatively close to each other. For this reason, defining a traffic sign dataset that contemplates the aforementioned problem, conducts our work to make an important contribution for intelligent vehicles.

Simultaneously, by training several state of the art CNN models, we showed that Deep CNNs are not required to solve traffic sign classification. Instead, techniques like image preprocessing and data-augmentation are used to improve classification accuracy. Accuracies of 99.37% and 98.99% were the best results obtained for training the GTSRB and our European dataset respectively with the modified CNN 8-layers model.

However, the classes based on text and with very different aspect ratios (most belonging to Informative category) were the most challenging ones to learn. This problem comes from

the input definition of the CNNs since they require a fixed size (most of the time squared). In consequence, information might be discarded when downscaling the image, or distorted from the original input.

The nature of Information category classes makes them hard to learn since, as they contain text and vary depending on the country, the appearance is always different. This category was proposed in the dataset with the aim to allow other researchers to choose the signs they are interested in and provide data to solve other complex tasks like text recognition and interpretation. At the same time, it can open new horizons to interpret correctly the traffic signs when additional panels (Informative subcategory) are located below a traffic sign.

In future work we intent to take into account the class imbalance problem to improve recognition accuracy. In this regard we will: 1) take into account more transformations for data-augmentation as performed in [15], 2) generate data simulating night scenarios with different levels of brightness changes and noise additions 3) apply class independent transformations like horizontal flip to signs which allow changing one class to another and 4) vertical flip to signs that do not change meaning. Besides the classification, we will also attempt to 5) detect all traffic signs taking into account all kinds of shapes and colors to provide a complete system for autonomous vehicles.

REFERENCES

[1] C. Grigorescu and N. Petkov, "Distance sets for shape filters and shape recognition," *IEEE Trans. Image Process.*, vol. 12, no. 10, pp. 1274–1286, Oct. 2003.

[2] S. Šegvic *et al.*, "A computer vision assisted geoinformation inventory for traffic infrastructure," in *Proc. 13th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Sep. 2010, pp. 66–73.

[3] F. Larsson and M. Felsberg, "Using Fourier descriptors and spatial models for traffic sign recognition," in *Scandinavian Conference on Image Analysis*. Berlin, Germany: Springer, 2011, pp. 238–249.

[4] R. Timofte and L. Van Gool, "Sparse representation based projections," in *Proc. 22nd Brit. Mach. Vis. Conf.-BMVC*, 2011, pp. 1–61.

[5] N. Paparoditis *et al.*, "Stereopolis II: A multi-purpose and multi-sensor 3D mobile mapping system for street visualisation and 3D metrology," *Revue Française Photogramm. Télédétection*, vol. 200, no. 1, pp. 69–79, 2012.

[6] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Netw.*, vol. 32, pp. 323–332, Aug. 2012.

[7] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," *Mach. Vis. Appl.*, vol. 25, no. 3, pp. 633–647, Apr. 2014.

[8] M. M. Lau, K. H. Lim, and A. A. Gopalai, "Malaysia traffic sign recognition with convolutional neural network," in *Proc. IEEE Int. Conf. Digit. Signal Process. (DSP)*, Jul. 2015, pp. 1006–1010.

[9] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1484–1497, Dec. 2012.

[10] S. Jung, U. Lee, J. Jung, and D. H. Shim, "Real-time traffic sign recognition system with deep convolutional neural network," in *Proc. 13th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, 2016, pp. 1–4.

[11] S. B. Wali, M. A. Hannan, A. Hussain, and S. A. Samad, "Comparative survey on traffic sign detection and recognition: A review," in *Przegląd Elektrotechniczny*. 2015.

[12] Economic Commission for Europe-Inland Tansport Committee, "Convention on road signs and signals," *United Nations Treaty Ser.*, vol. 1091, p. 3, Nov. 1968.

[13] Y. Saadna and A. Behloul, "An overview of traffic sign detection and classification methods," *Int. J. Multimedia Inf. Retr.*, vol. 6, no. 3, pp. 193–210, 2017.

[14] H. Fleyeh and M. Dougherty, "Road and traffic sign detection and recognition," in *Proc. 16th Mini-EURO Conf. 10th Meeting EWGT*, 2005, pp. 644–653.

[15] H. H. Aghdam, E. J. Heravi, and D. Puig, "A practical and highly optimized convolutional neural network for classifying traffic signs in real-time," *Int. J. Comput. Vis.*, vol. 122, no. 2, pp. 246–269, 2017.

[16] Y. Yang, H. Luo, H. Xu, and F. Wu, "Towards real-time traffic sign detection and classification," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 7, pp. 2022–2031, Jul. 2016.

[17] S. M. Bascón, J. A. Rodríguez, S. L. Arroyo, A. F. Caballero, and F. López-Ferreras, "An optimization on pictogram identification for the road-sign recognition task using SVMs," *Comput. Vis. Image Understand.*, vol. 114, no. 3, pp. 373–383, 2010.

[18] G. Piccioli, E. De. Micheli, P. Parodi, and M. Campani, "Robust method for road sign detection and recognition," *Image Vis. Comput.*, vol. 14, no. 3, pp. 209–223, 1996.

[19] X. W. Gao, L. Podladchikova, D. Shaposhnikov, K. Hong, and N. Shevtsova, "Recognition of traffic signs based on their colour and shape features extracted using human vision models," *J. Vis. Commun. Image Represent.*, vol. 17, no. 4, pp. 675–685, Aug. 2006.

[20] A. Ruta, Y. Li, and X. Liu, "Robust class similarity measure for traffic sign recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 11, no. 4, pp. 846–855, Dec. 2010.

[21] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The german traffic sign detection benchmark," in *Proc. IEEE Int. Joint Conf. Neural Netw. (IJCNN)*, Aug. 2013, pp. 1–8.

[22] Z. Abedin, P. Dhar, M. K. Hossenand, and K. Deb, "Traffic sign detection and recognition using fuzzy segmentation approach and artificial neural network classifier respectively," in *Proc. IEEE Int. Conf. Elect., Comput. Commun. Eng. (ECCE)*, Feb. 2017, pp. 518–523.

[23] K. T. Islam and R. G. Raj, "Real-time (vision-based) road sign recognition using an artificial neural network," *Sensors*, vol. 17, no. 4, p. 853, 2017.

[24] Y. Li, A. Møgelmose, and M. M. Trivedi, "Pushing the 'speed limit': High-accuracy US traffic sign recognition with convolutional neural networks," *IEEE Trans. Intell. Vehicles*, vol. 1, no. 2, pp. 167–176, Jun. 2016.

[25] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[26] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ.Toronto, Toronto, ON, Canada, Tech. Rep., 2009.

[27] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[28] D. Cireşan, U. Meier, and J. Schmidhuber. (Feb. 2012). "Multi-column deep neural networks for image classification." [Online]. Available: https://arxiv.org/abs/1202.2745

[29] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. (Dec. 2013). "OverFeat: Integrated recognition, localization and detection using convolutional networks." [Online]. Available: https://arxiv.org/abs/1312.6229

[30] J. Jin, K. Fu, and C. Zhang, "Traffic sign recognition with hinge loss trained convolutional neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1991–2000, Oct. 2014.

[31] J. Li and Z. Wang, "Real-time traffic sign recognition based on efficient CNNs in the wild," *IEEE Trans. Intell. Transp. Syst.*, to be published.

[32] R. Belaroussi, P. Foucher, J.-P. Tarel, B. Soheilian, P. Charbonnier, and N. Paparoditis, "Road sign detection in images: A case study," in *Proc. 20th IEEE Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2010, pp. 484–488.

[33] H. Luo, Y. Yang, B. Tong, F. Wu, and B. Fan, "Traffic sign recognition using a multi-task convolutional neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 4, pp. 1100–1111, Apr. 2018.

[34] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1026–1034.
[35] V. Nair and G. E. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, 2010, pp. 807–814.
[36] S. Ioffe and C. Szegedy. (Mar. 2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167
[37] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[38] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.
[39] K. Simonyan and A. Zisserman. (2014). "Very deep convolutional networks for large-scale image recognition." [Online]. Available: https://arxiv.org/abs/1409.1556
[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[41] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "CNN features off-the-shelf: An astounding baseline for recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2014, pp. 806–813.
[42] S. Chilamkurthy. (2017). *Keras Tutorial—Traffic Sign Recognition*. [Online]. Available: http://chsasank.github.io/keras-tutorial.html
[43] D. P. Kingma and J. Ba. (Dec. 2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980
[44] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

CITLALLI GÁMEZ SERNA received the B.S. degree in information and communication technologies from the Instituto Tecnológico y de Estudios Superiores de Monterrey, Mexico, in 2011, and the M.S. degree in color in informatics and media technology (Erasmus program) from Université Jean Monnet, France, and the Gjøvik University College, Norway, respectively, in 2014. She is currently pursuing the Ph.D. degree in informatics and automation with the Systems and Transportations Laboratory, University of Technology of Belfort-Montbéliard, France. Her research is focused on path tracking for intelligent vehicles covering deep learning techniques for the environment perception and localization together with the control part.

YASSINE RUICHEK received the Ph.D. degree in control and computer engineering and the Habilitation à Diriger des Recherches degree in physic science from the University of Lille, France, in 1997 and 2005, respectively. Since 2007, he has been a Full Professor with the University of Technology of Belfort-Montbéliard. His research interests are concerned with multisensory data-based perception and localization, including computer vision, pattern recognition and classification, machine learning, and data fusion, with applications to intelligent transportation systems and video surveillance.