GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Towards Explainable AI Systems for Traffic Sign Recognition and Deployment in a Simulated Environment

Presented by
Pascal Fischer & Alen Smajic

with support of the
AI Systems Engineering Lab

Motivation and Introduction

- Recent advances in Machine Learning pushing the boundaries of real world AI applications
- Numerous challenges for our society and the system engineers
- Ethical AI, privacy and transparency are becoming increasingly relevant

- First attempt at tackling the problem of traffic sign recognition as systems engineers
- Focus on the development of an explainable and transparent system with the use of concept whitening layers
- Introduction of a simulation used for model deployment, performance benchmarking, data generation and further domain model research

1. User View

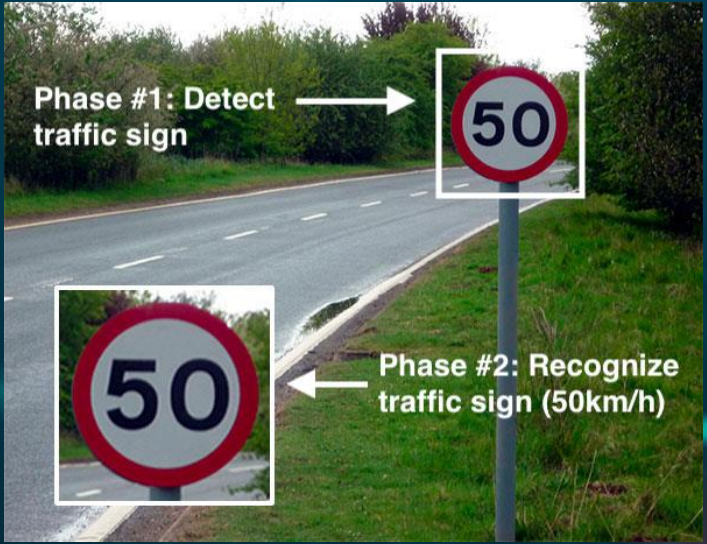
1.1 CONTEXT

- Traffic signs are part of the road infrastruture
- Traffic sign locations and positioning on the road
- Traffic sign shapes, colors and materials
- Outdoor environment and its influence on visibility
- Accidents, vandalism and weather influence on traffic signs

1. User View

1.2 TASK

- Recognition of traffic signs in real-time
- Detect and classify signs in various conditions and environments
- Focus on transparency and explainability

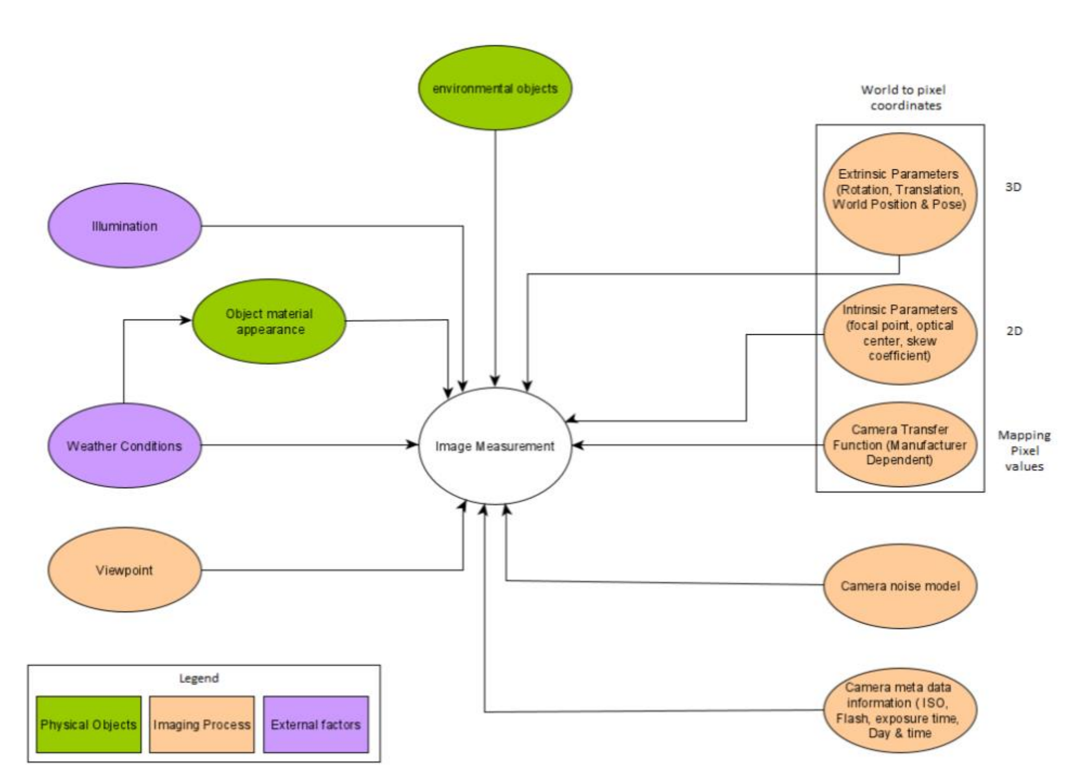


1. User View

1.3 PERFORMANCE CONSTRAINTS

- Real-time performance
- High prediction performance with high accuracy
- High recall for the detection and high accuracy for the classification
- High detection rate of small objects (on great distances)

2. Modeling View

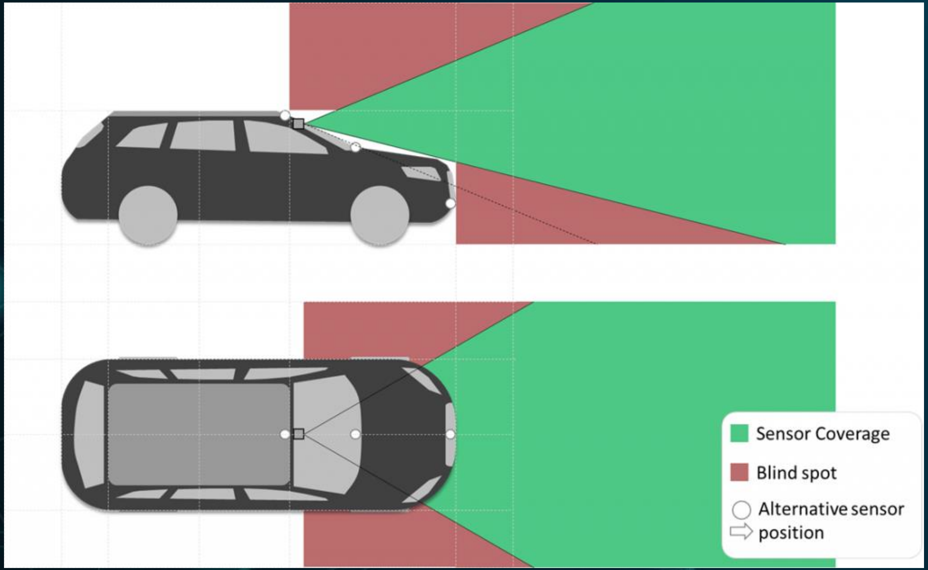


2.1 CONTEXTUAL WORLD MODEL

- Centered around the image measurement

2. Modeling View

2.1 CONTEXTUAL WORLD MODEL



CAMERA SENSOR MODEL

- Mounted on the roof, facing to the front
- Big field of view, reducing blind spots
- Robustness to camera movement and shaking
- Good performance in low illumination conditions
- Clear high resolution images at any time

2. Modeling View

2.1 CONTEXTUAL WORLD MODEL

TRAFFIC SIGN APPEARANCE

43 Traffic Signs from German Traffic Sign Recognition Benchmark

Traffic sign font

Linear-Antiqua family according to DIN 1451-2

Traffic sign sizes

traffic sign shapes	size 1 (70 %)	size 2 (100 %)	size 3 (125 or 140 %)
circles (diameter)	420	600	750 (125 %)
triangle (side length)	630	900	1260 (140 %)
square (side length)	420	600	840 (140 %)
rectangle (W x H)	630 x 420	900 x 600	1260 x 840 (140 %)

Speed limit for size class

speed limit (km/h) on circular shape	speed limit (km/h) on other shapes	size
0 to 20	20 to less than 50	1
20 to 80	50 to 100	2
more than 80	more than 100	3

2. Modeling View

2.1 CONTEXTUAL WORLD MODEL

TRAFFIC SIGN PLACEMENT

- Maximum of three signs on a single post (two if regulatory signs)
- Placed in general on clearly visible locations on the right side of the road
- may be placed on the left alone or above the road alone, or on both sides
- may be curved in such a way that they can also be seen from the side
- lower edge should be 2m above street level, 2.20m above cycle paths, 4.50m on gantries, 0.60m on islands and traffic dividers
- must not be placed within the roadway. The lateral distance from it should be 0.50m within built-up areas and never less than 0.30m. For outside built-up areas it should be 1.50m

2. Modeling View

2.1 CONTEXTUAL WORLD MODEL

TRAFFIC SIGN COLORS

- The colors used for traffic signs are regulated by the DIN 6171
- The color charts „RAL-F 7 Reflexfarben“ and „RAL-F 81 Farben im Straßenverkehr“ show the colors of traffic signs in new condition
- The color shades specified in the RAL-F 7 special color series are:

  RAL 2006, RAL 3019, RAL 3030, RAL 5016, RAL 6030, RAL 8026, RAL 9014, RAL 9019

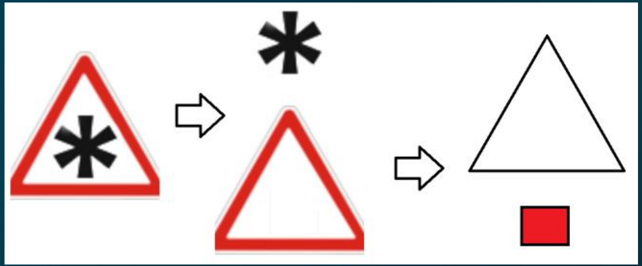
- The color shades specified in the RAL-F 81 special color series are:

  RAL 1023 (yellow), RAL 2009 (orange), RAL 3020 (red), RAL 4006 (purple), RAL 5017 (blue), RAL 6024 (green), RAL 7042 (gray A), RAL 7043 (gray B), RAL 9016 (white), RAL 9017 (black)

2. Modeling View

2.1 CONTEXTUAL WORLD MODEL

CONCEPTUAL DECOMPOSITION OF TRAFFIC SIGNS



Concepts consists of further concepts

„part of" hierarchy

Conceptual decomposition is ambigous

Partonomy

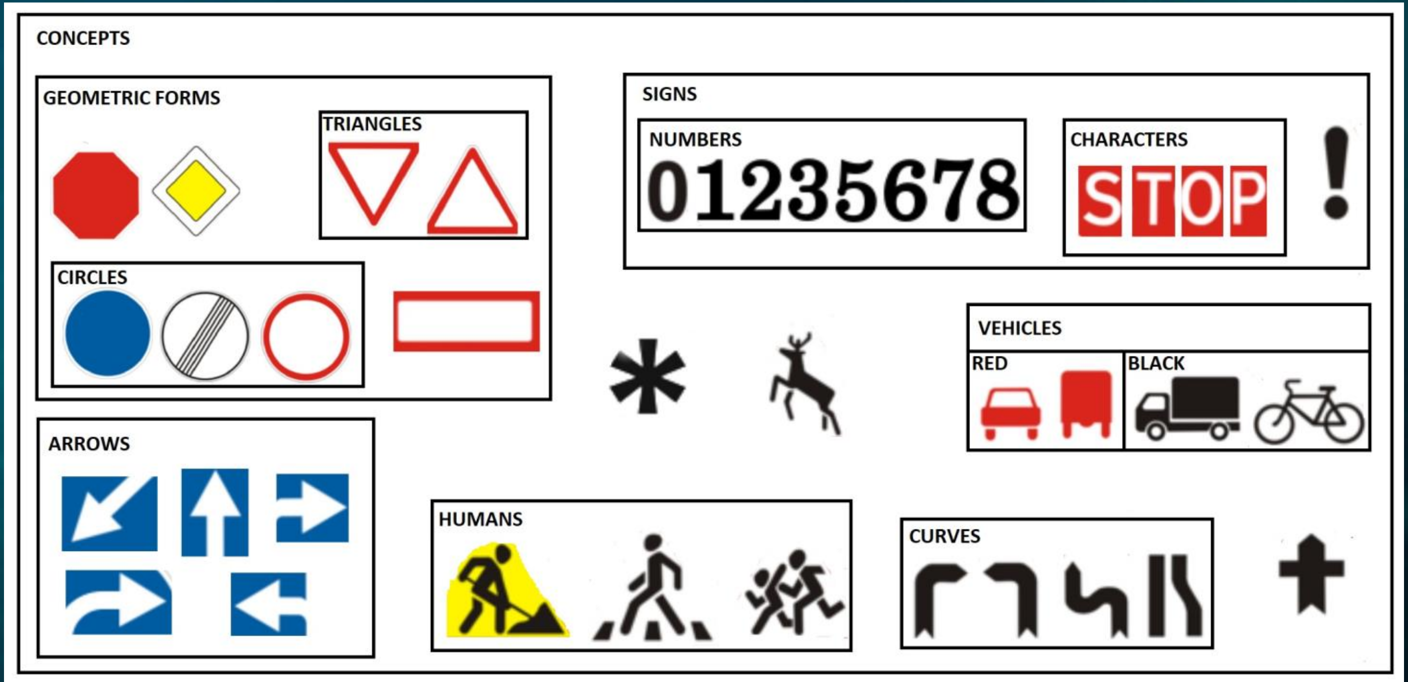
Decomposition could not be clearly determined

Depends on the observer

2. Modeling View

2.1 CONTEXTUAL WORLD MODEL

CONCEPTUAL DECOMPOSITION MODEL



2. Modeling View

2.1 CONTEXTUAL WORLD MODEL

PROBABILISTIC ENVIRONMENTAL WORLD MODEL

WEATHER	
Clear / Sunny	23 %
Overcast	45 %
Foggy	15 %
Rainy	23 %
Snowy	0.8 %
Thunderstorm	7 %

DAYTIME	
Day	51 %
Night	49 %

ILLUMINATION	
Sun illumination	51 %
Own car illumination	100 %
Other cars illumination	99 % *
Street lamps illumination	49 %
Other illumination sources	/

OBJECTS INSIDE THE SCENE	
People	50 % *
Animals	2 % *
Trees	90 % *
Other vegetation	95 % *
Vehicles	99 % *
Buildings	75 % *
Road objects (except signs)	95 % *
Traffic signs	90 % *

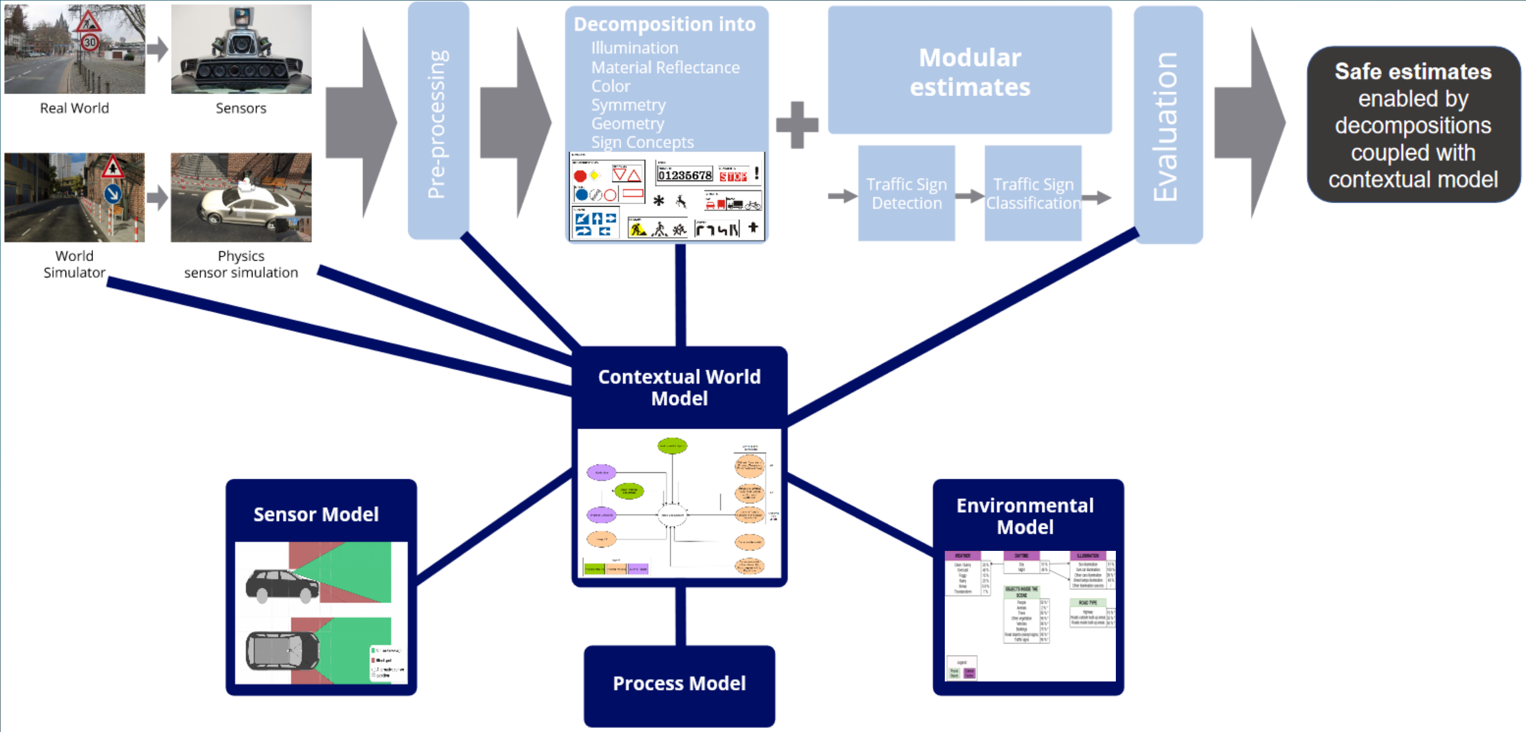
ROAD TYPE	
Highway	15 % *
Roads outside built-up areas	35 % *
Roads inside built-up areas	50 % *

Legend

Physial Objects	External Factors

2. Modeleing View

2.2 SYSTEM DESIGN



2. Modeling View

2.3 PRIVACY AND TRANSPARENCY BY DESIGN

- Protect the privacy of each individual affected by our system
- No video storage component => no privacy concerns
- Use camera sensor only for real-time traffic sign recognition

- Transparency by deploying concept whitening layers within our deep learning models

3. Implementation View

3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES

Selecting Haar-like features

Creating an integral image

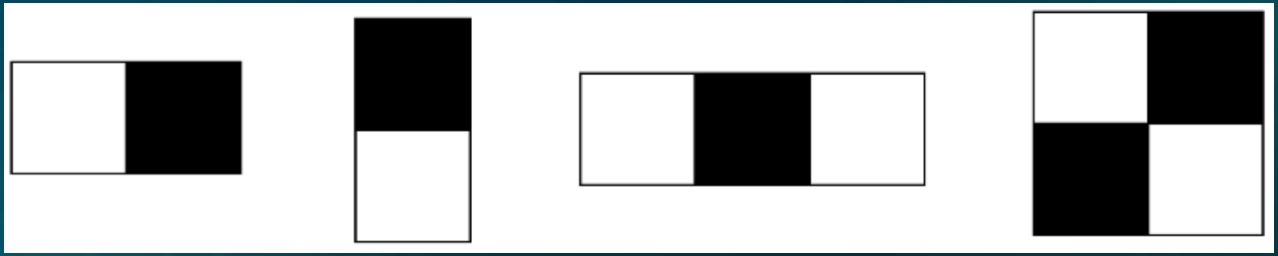
Running Adaboost training

Creating classifier cascades

3. Implementation View

3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES

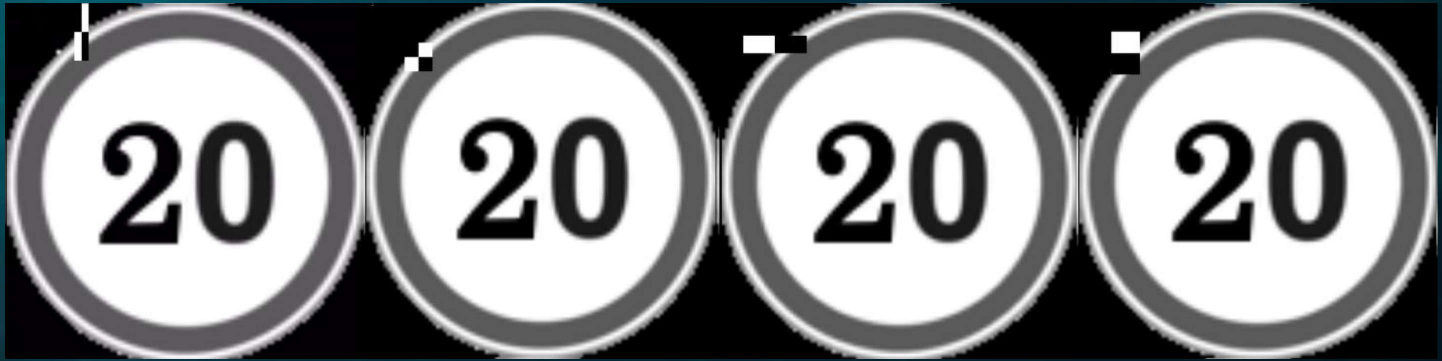
HAAR-LIKE FEATURES



Edge-features

Line-features

Four-sided-features



3. Implementation View

3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES

INTEGRAL IMAGE

	x_0	x_1	x_2	x_3
y_0	1	12	45	10
y_1	6	5	11	4
y_2	3	7	10	8
y_3	5	9	4	7

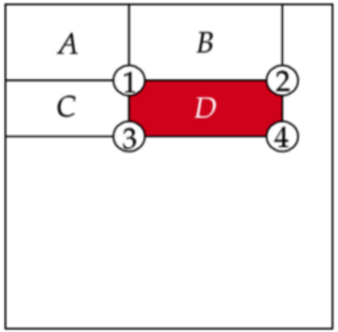
Original Image
(Grayscale)

$$ii(x,y) = \sum_{x' \leqslant x,\, y' \leqslant y} i(x',y')$$

i – original image
ii – integral image

	x_0	x_1	x_2	x_3
y_0	1	13	58	68
y_1	7	24	80	94
y_2	10	34	100	122
y_3	15	48	118	147

Integral Image



$$D = 4 - 2 - 3 + 1$$
$$= 4 + 1 - (2 + 3)$$

3. Implementation View

3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES

- Given example images $(x_1, y_1), \ldots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \ldots, T$:
  1. Normalize the weights,
     $$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$$
     so that w_t is a probability distribution.
  2. For each feature, j, train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t, $\epsilon_j = \sum_i w_i \, |h_j(x_i) - y_i|$.
  3. Choose the classifier, h_t, with the lowest error ϵ_t.
  4. Update the weights:
     $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$
     where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.
- The final strong classifier is:
  $$h(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$
  where $\alpha_t = \log \frac{1}{\beta_t}$

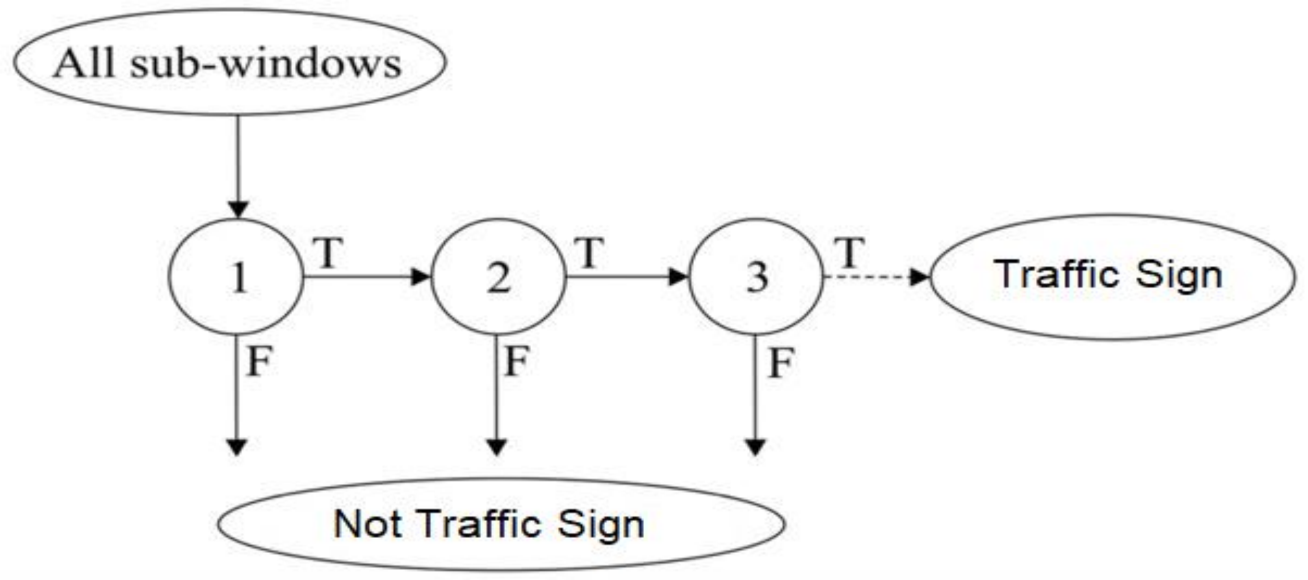
ADABOOST

- ~160,000 possible features in a 24x24 detector window
- Identify the best features
- Each feature is a weak learner
- Checks performance of all classifiers that are supplied to a feature
- Classification by strong response as positive, weak as negative
- Result is a strong classifier -> boosted classifier

3. Implementation View

3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES

CASCADE CLASSIFIER



3. Implementation View

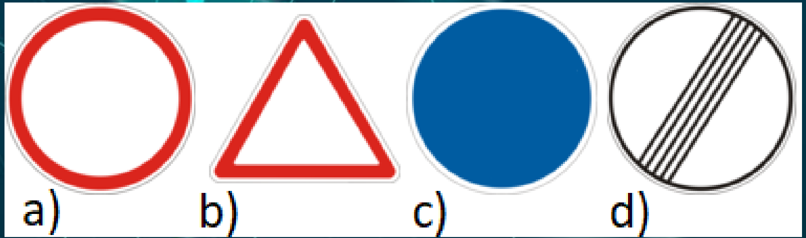
3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES

DATASET

- Positive examples from rebalanced GTSRB
- Negative examples from simulation data, GTSDB, and downloaded images
- Preprocessing with Spatial Transformer Network STN

3. Implementation View

3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES



TRAINING

- 4 different classifiers for classes :
- a) white circle red border
- b) white triangle red border
- c) blue circle
- d) white circle with black diagonal lines
- Minimum hit rate of 99%
- Maximum false alarm rate of 30%
- 1000 positive and 2000 negative examples
- train only on main color channel and grayscale

3. Implementation View

3.1 TRAFFIC SIGN DETECTION USING VIOLA & JONES

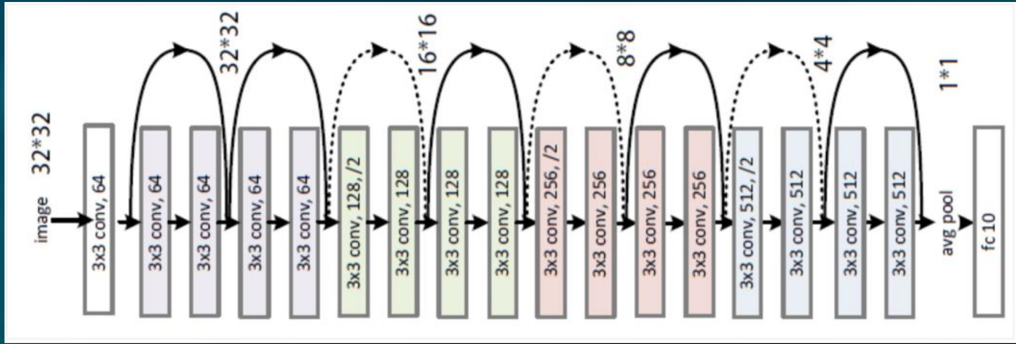
class	color	recall
red circle	grayscale	79
	red	71
red triangle	grayscale	88
	red	81
blue circle	grayscale	48
	blue	64
white slashes	grayscale	83

VALIDATION

- Test dataset from GTSDB
- 600 images
- ~1000 bounding boxes
- blue circle gets better result on blue color channel

3. Implementation View

3.2 TRAFFIC SIGN CLASSIFICATION



ResNet18

Concept Whitening

3. Implementation View

3.2 TRAFFIC SIGN CLASSIFICATION

DATASET - GTSRB

- German Traffic signs
- 50.000 collected images
- 43 classes
- Different weather conditions and illumination
- Variations of distance, partial occlusion and rotation
- 12.500 images in testset

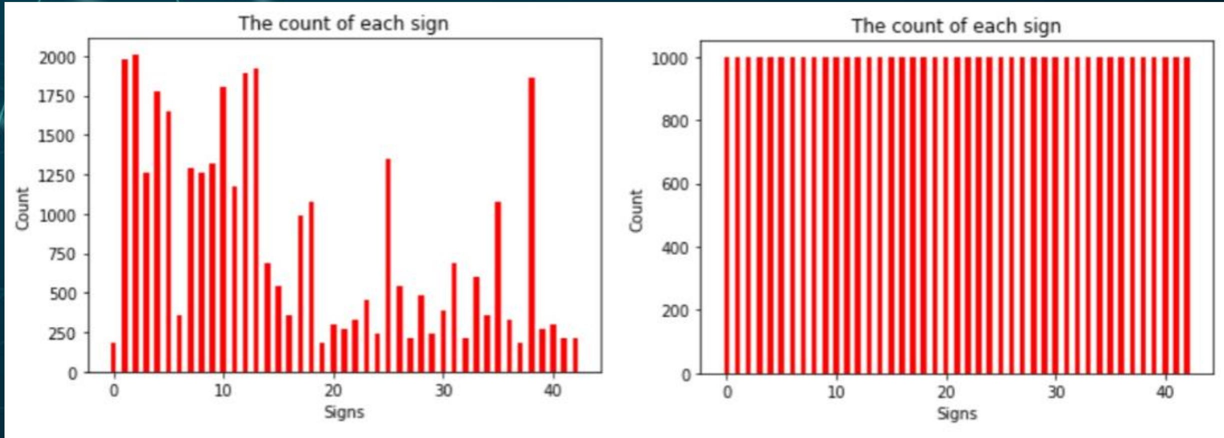
3. Implementation View

3.2 TRAFFIC SIGN CLASSIFICATION

DATA PREPROCESSING

- Choose 1000 random images of each class with at least 1000 images

- Add for each image of a class with less than 1000 images an additional rotated version with angle between [-20, 20]



(a) unbalanced dataset (b) preprocessed balanced dataset

3. Implementation View

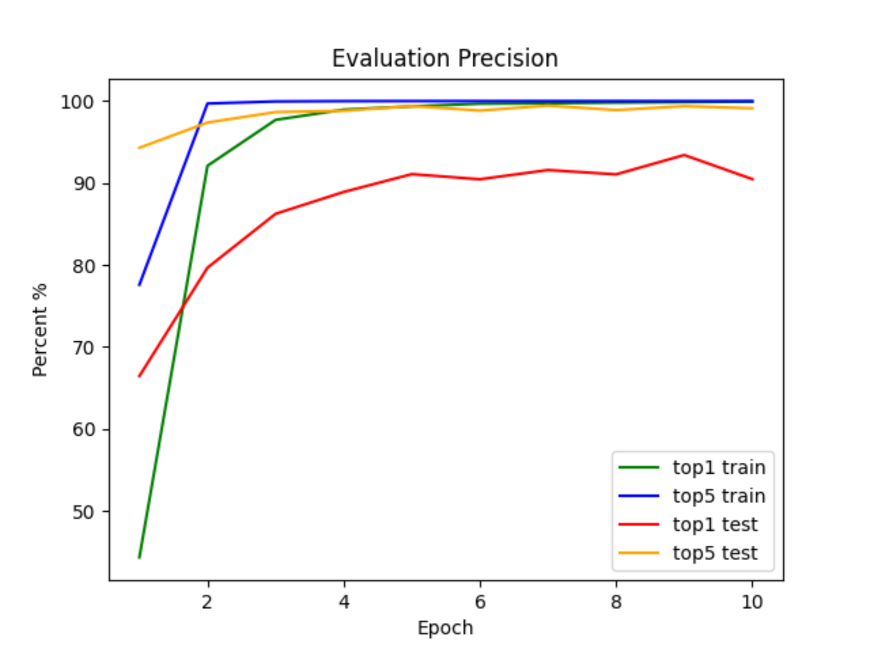
3.2 TRAFFIC SIGN CLASSIFICATION

TRAINING

Background Class	Epochs	Learning Rate	Momentum	Optimizer	Loss Function
1000 images from negatives of object detection dataset	10	1e-2	9e-1	Stochastic Gradient Descent	Cross Entropy Loss

3. Implementation View

3.2 TRAFFIC SIGN CLASSIFICATION



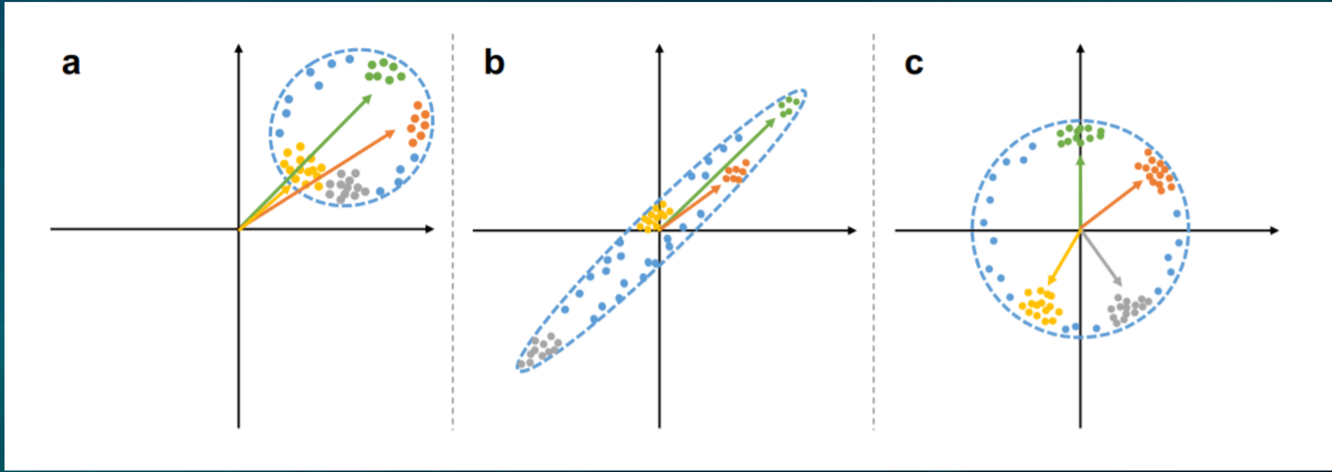
VALIDATION

- top1 and top5 precision for train- and testset

train		test	
top1	top5	top1	top5
99.873	99.998	93.413	99.375

3. Implementation View

3.3 CONCEPT WHITENING



data are not mean centered

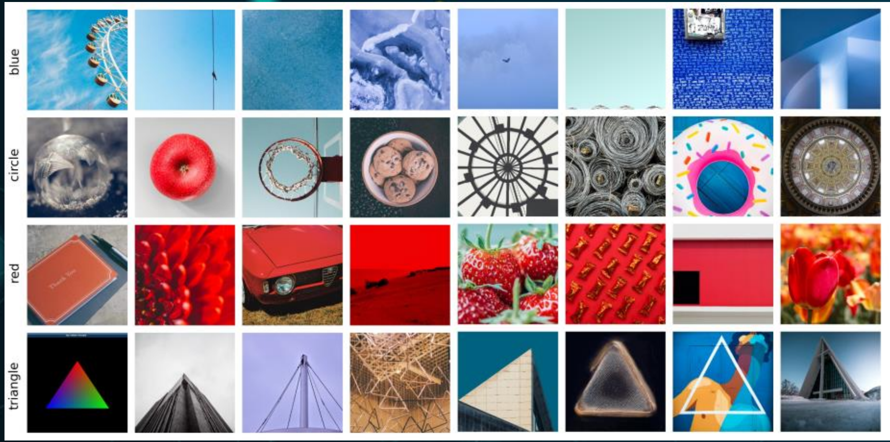
standardized data but not decorrelated

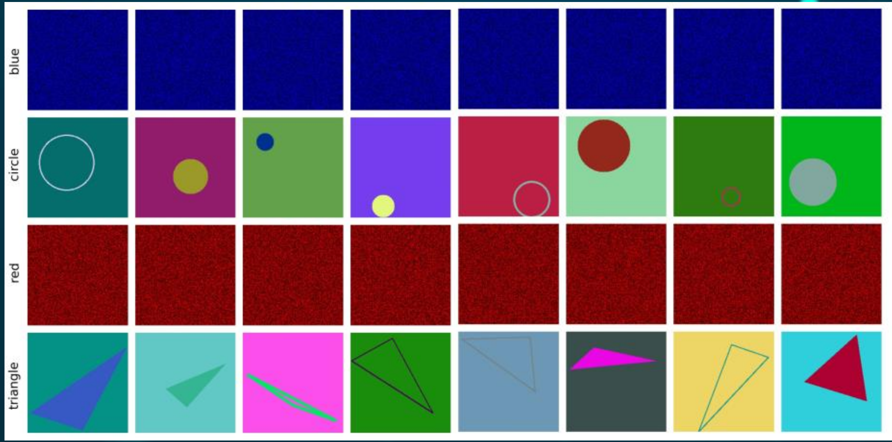
whitened data

3. Implementation View

3.3 CONCEPT WHITENING

CONCEPT DATASETS





Internet Collection
500 images per concept

Formular generated
1000 images per concept

3. Implementation View

3.3 CONCEPT WHITENING

TRAINING

5th residual block
Replace batch normalization layer in pretrained ResNet18

Same parameters
Training with the same parameters as without CW except learn rate of 5e-2

5 epochs
Train additional epochs on trained classification network

3. Implementation View

3.3 CONCEPT WHITENING

VALIDATION 1/2

dataset	concept	1 epoch	5 epochs
selfmade	blue	0.894	0.932
	circle	0.671	0.670
	red	0.610	0.959
	triangle	0.537	0.344
downloaded	blue	0.9735	0.970
	circle	0.684	0.658
	red	0.845	0.923
	triangle	0.709	0.582

Concept validation with AUC

dataset	epochs	top1	top5
without	9	93.413	99.375
selfmade	10	90.792	99.097
	14	93.674	99.367
downloaded	10	91.696	99.305
	14	92.352	98.971

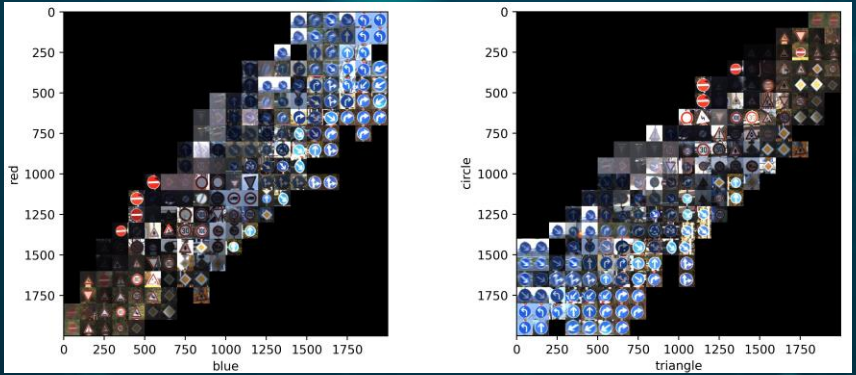
Traffic sign recognition validation with top1, top5 precision

3. Implementation View

3.3 CONCEPT WHITENING

VALIDATION 2/2

Top 10 activated images



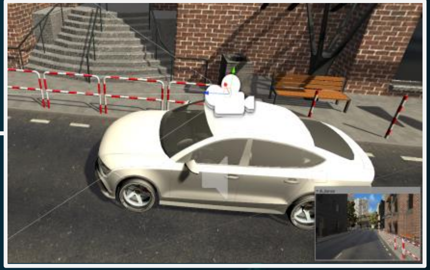
Two concepts subspace

3. Implementation View

3.4 SIMULATION

Environmental component

Car component



Traffic sign component

Traffic sign recognition system

3. Implementation View

3.4 SIMULATION

ENVIRONMENTAL COMPONENT

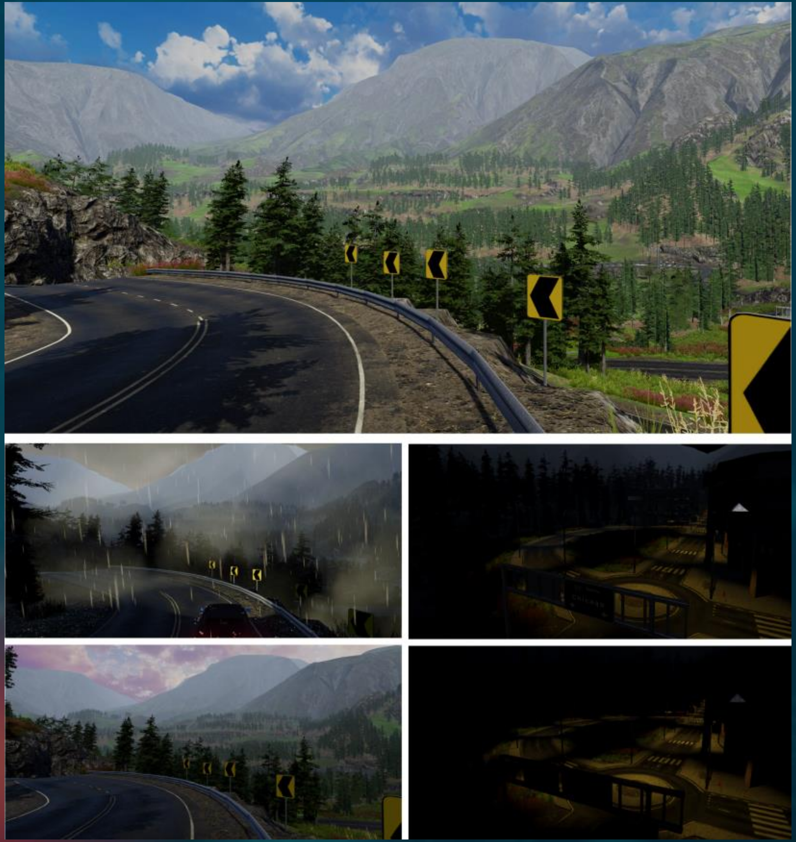
Windridge City simulation environment

Urban roads, rural area roads, highways and forest roads

Different road textures and floor markings

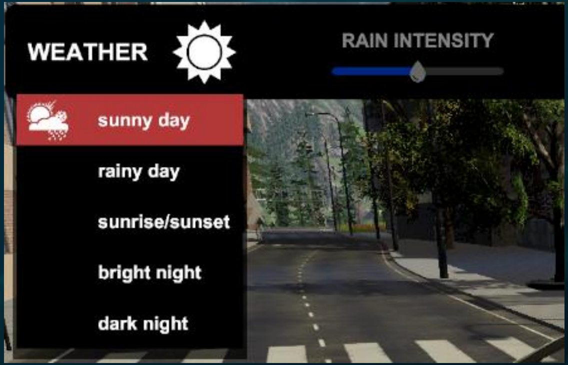
3. Implementation View

3.4 SIMULATION



ENVIRONMENTAL COMPONENT

- Weather system with 5 options:
  sunny day, rainy day, sunrise/sunset, bright night, dark night



3. Implementation View

3.4 SIMULATION

INTEGRATION OF THE CAR COMPONENT

- 3D car model
- Brake lights, active brake lights, low beam headlights, long beam headlights, turn signal lights
- Camera sensor mounted on the roof
- Wheel collider component
- 2 driving modes: autonomous driving and manual driving
- 5 different camera modes

3. Implementation View

3.4 SIMULATION

INTEGRATION OF THE TRAFFIC SIGN COMPONENT

- Manually generated 3D traffic sign models
- Traffic sign spawn function with the following parameters: traffic sign spawning frequency, frequency of double signs, frequency of rotation variance, frequency of occluded signs
- Active signs option

3. Implementation View

3.4 SIMULATION

INTEGRATION OF THE TRAFFIC SIGN COMPONENT

Spawn Frequency

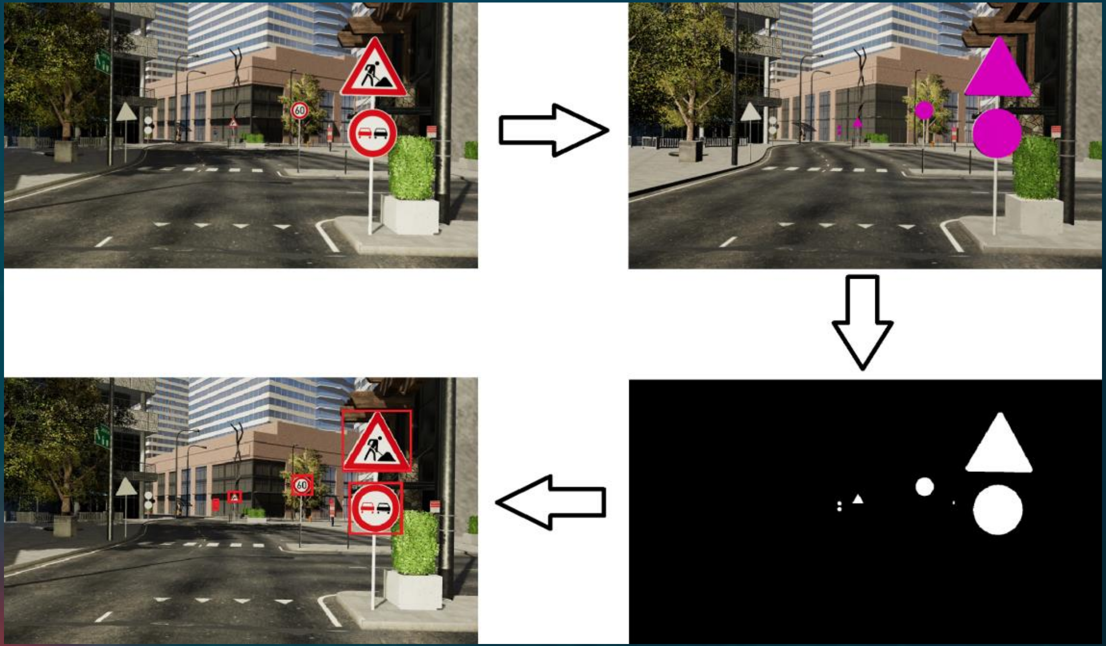
Double Signs Frequency

Frequency of Rotation Variance

Frequency of Occluded Signs

3. Implementation View

3.4 SIMULATION



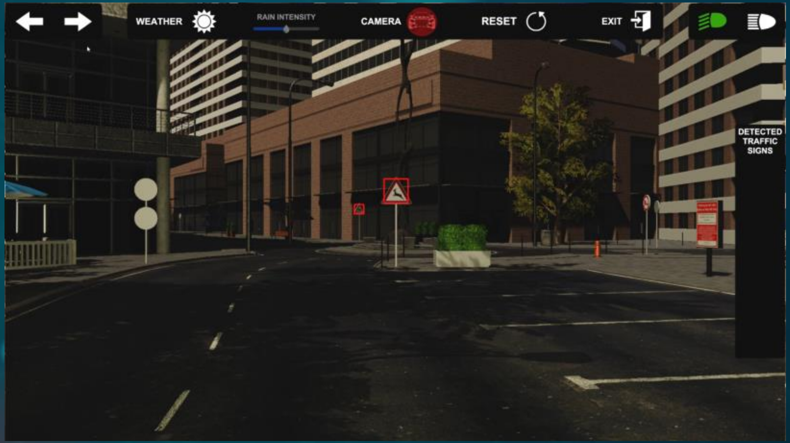
DATASET GENERATION

- Stops the scene every few seconds to make 2 screenshots
- First screenshot is the sensor input
- Second screenshot is the ground truth label (pink texture on traffic signs)
- Threshold the color and use the contours function to extract the bounding box labels

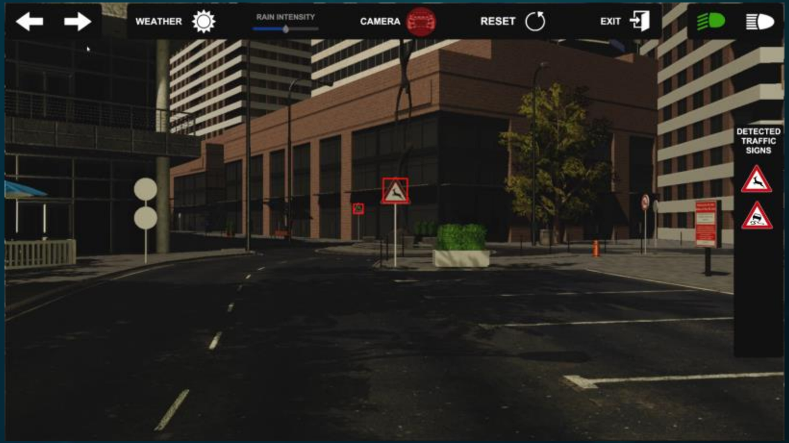
3. Implementation View

3.4 SIMULATION

INTEGRATION OF THE TRAFFIC SIGN RECOGNITION SYSTEM

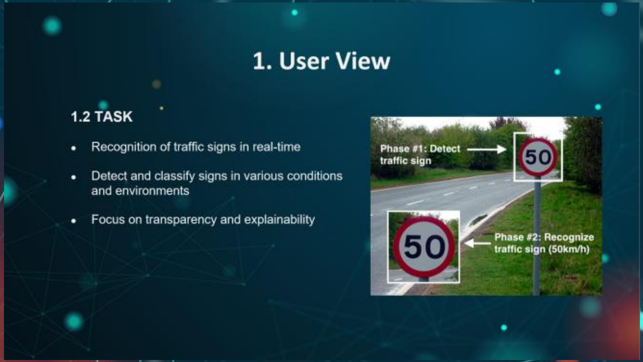


Our current system



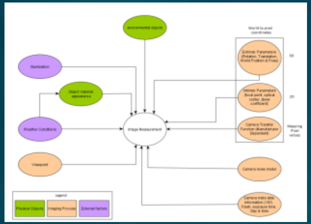
Visualisation of the future version of our system

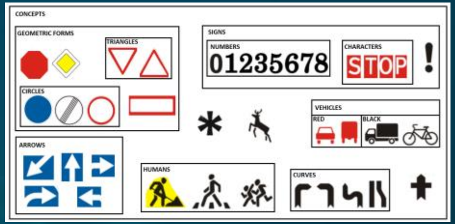
4. Conclusion and Future Work

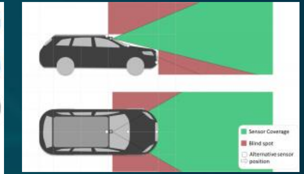


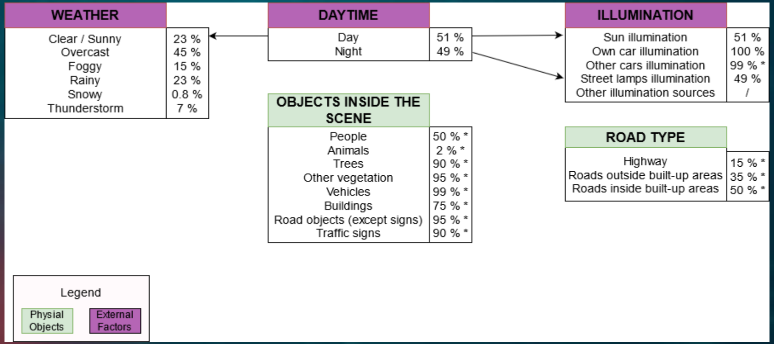
- We presented our first attempt at tackling the problem of traffic sign recognition as system engineers

4. Conclusion and Future Work



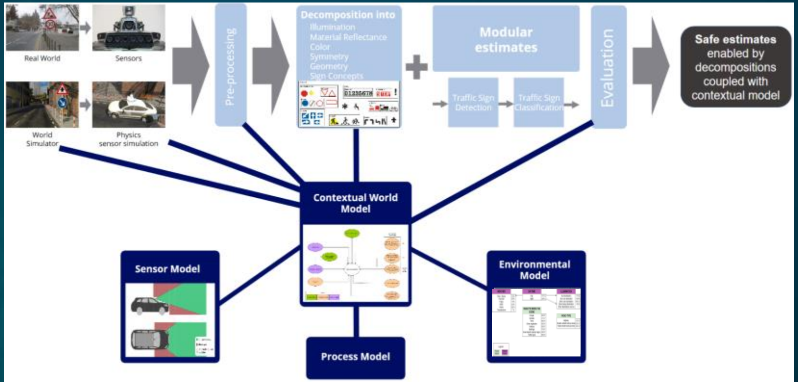


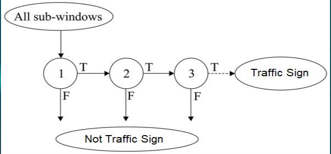


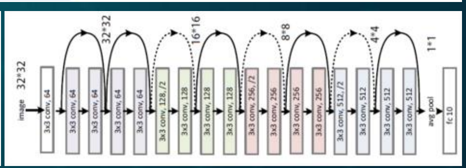


- Contextual world model
- Camera sensor model
- Try out new algorithms for the traffic sign detection
- Expand the amount of concept whitening layers within our networks
- Try out different image classification architectures

4. Conclusion and Future Work

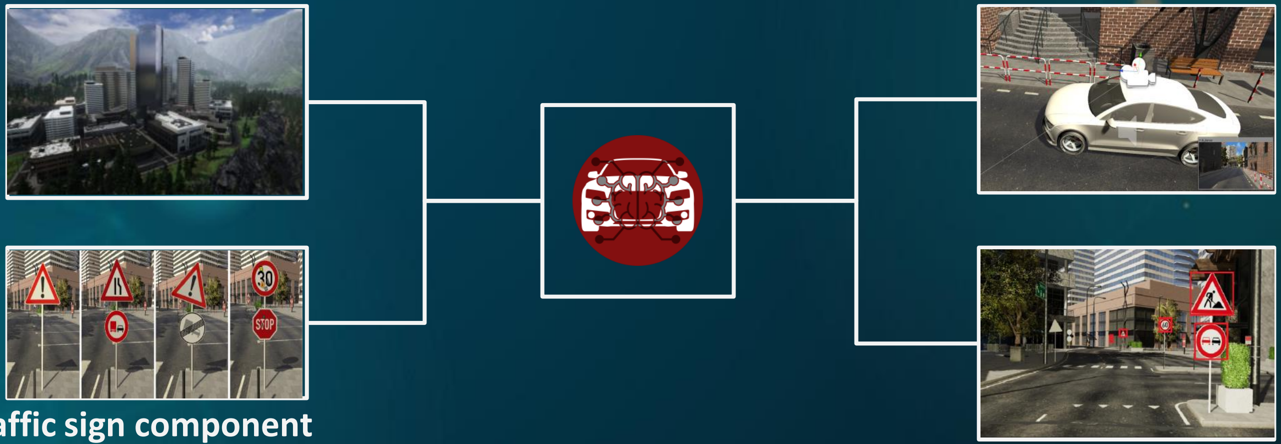






- Multistage pipeline system
- Future work => exploit the fact that our system consists of 2 prediction models
- Try out new algorithms for the traffic sign detection system
- Expand the amount of concept whitening layers within our networks
- Try out different image classification architectures

4. Conclusion and Future Work



- We presented a simulation for the task of traffic sign recognition
- Future work => Further development of the simulation components with respect to our probabilistic environmental world model
- The simulation as a benchmarking tool

Thank You

For Your Attention

Presented by
Pascal Fischer & Alen Smajic

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

with support of the
AI Systems Engineering Lab