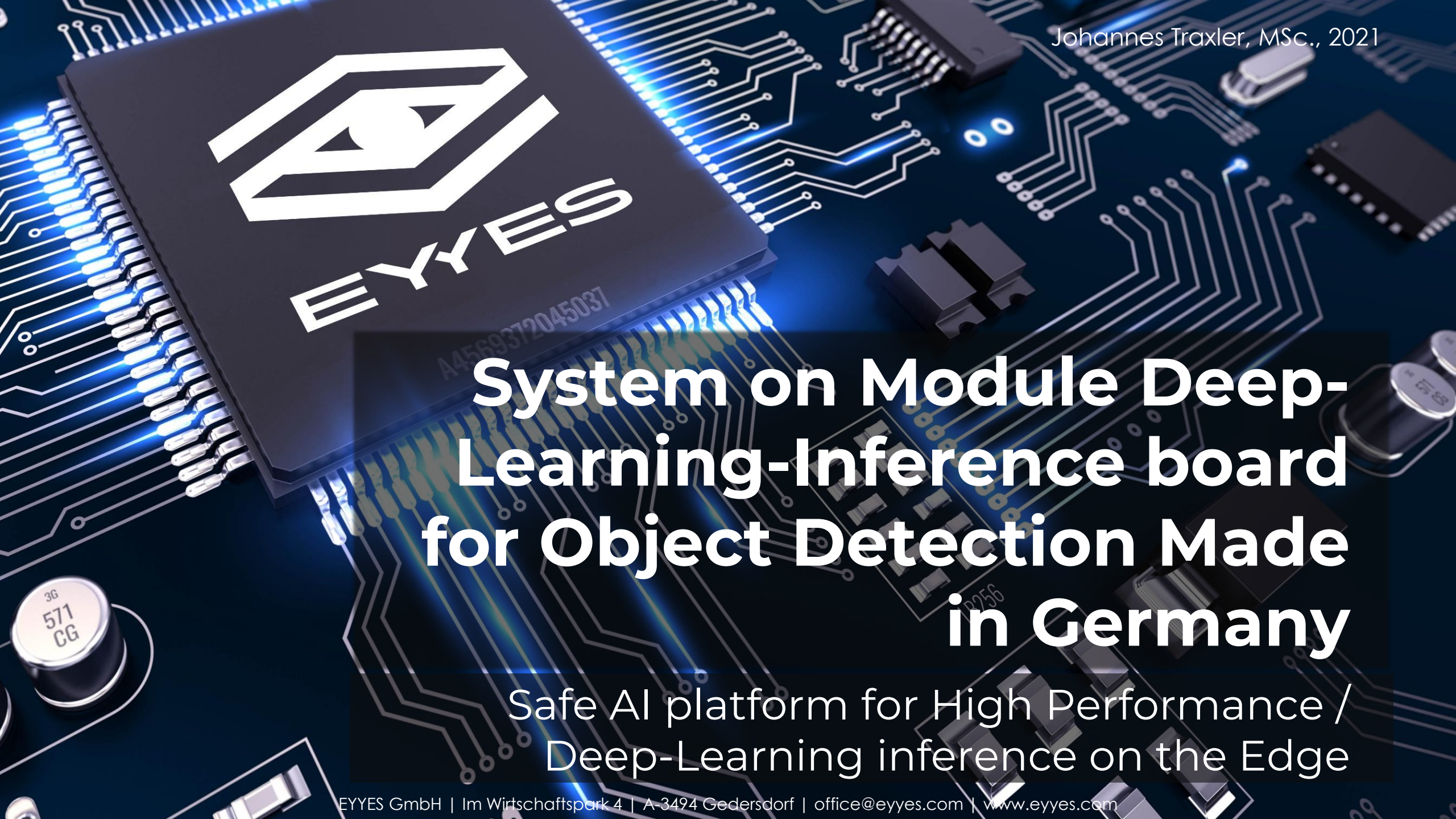


Johannes Traxler, MSc., 2021

System on Module Deep-Learning-Inference board for Object Detection Made in Germany

Safe AI platform for High Performance / Deep-Learning inference on the Edge

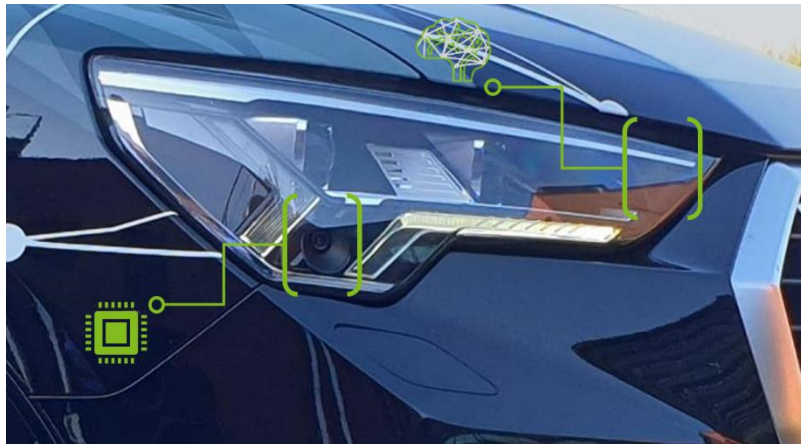
EYYES GmbH | Im Wirtschaftspark 4 | A-3494 Gedersdorf | office@eyyes.com | www.eyyes.com

EYYES - We make machines see

Technology leader for Safe Artificial Intelligence

EYYES stands for:

- … make our world safer with artificial intelligent
- … experts in computer vision, machine learning and embedded systems
- … camera and sensing technology
- … soft- and hardware development made in Europe
- … technological, tailor-made solutions at the highest level through lead, competence and innovation
- … strong relationship to leading Europe machine learning research organizations



we make machines see

Sites

KREMS AN DER DONAU / Austria

Headquarter

- General Management
- Sales & Marketing
- Project Management & Execution
- Assembling / Test Field
- Procurement
- Research & Development

AACHEN / Germany

Competence Center Software Engineering

- Development Software
- Development Algorithmic
- Development Artificial Intelligence & Machine Learning (Deep Learning)
- Research & Development

FREITAL (DRESDEN) / Germany

Competence Center Embedded Systems

- Development Embedded Systems
- Service & Repair
- Electronics Laboratory & Certifications
- Safety Engineering
- Research & Development

EYYES

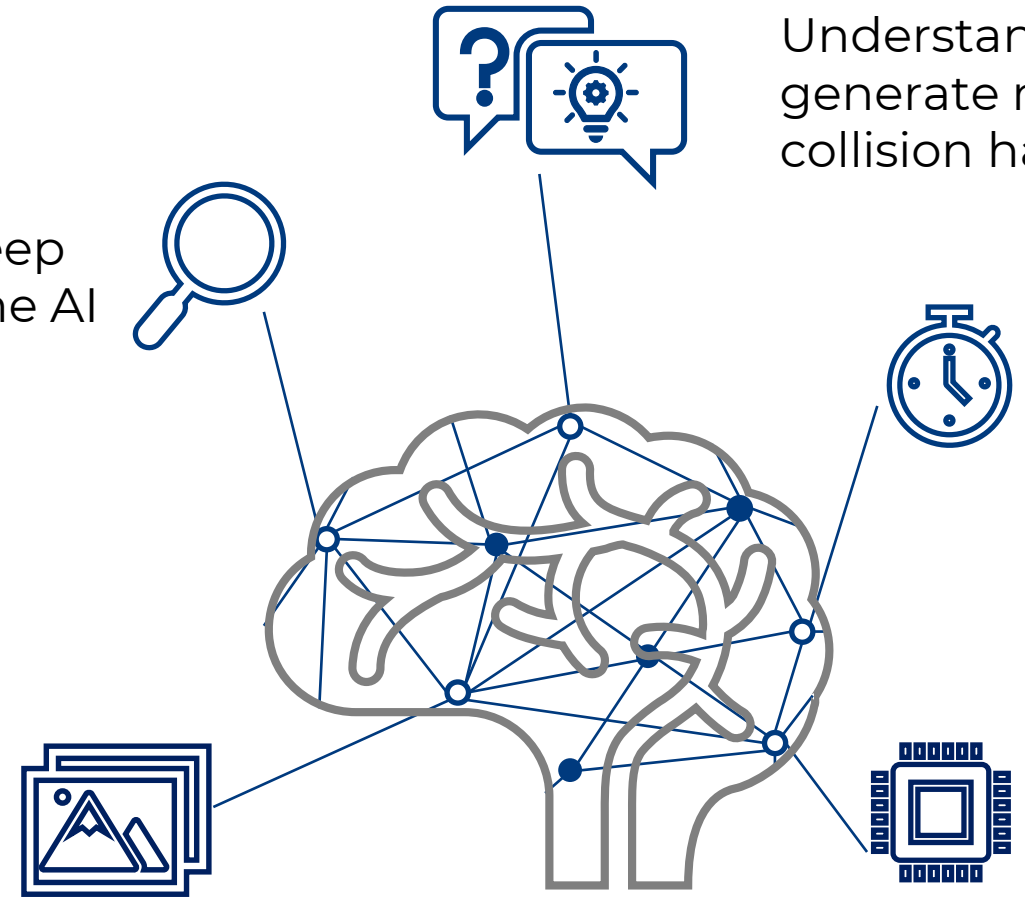
we make machines see

Smart solutions for safe mobility

with Deep learning

UNDERSTANDING
Understand complex situations, generate motion predictions, detect collision hazards

ANALYSIS
Classify objects by deep neural networks of the AI detection modules

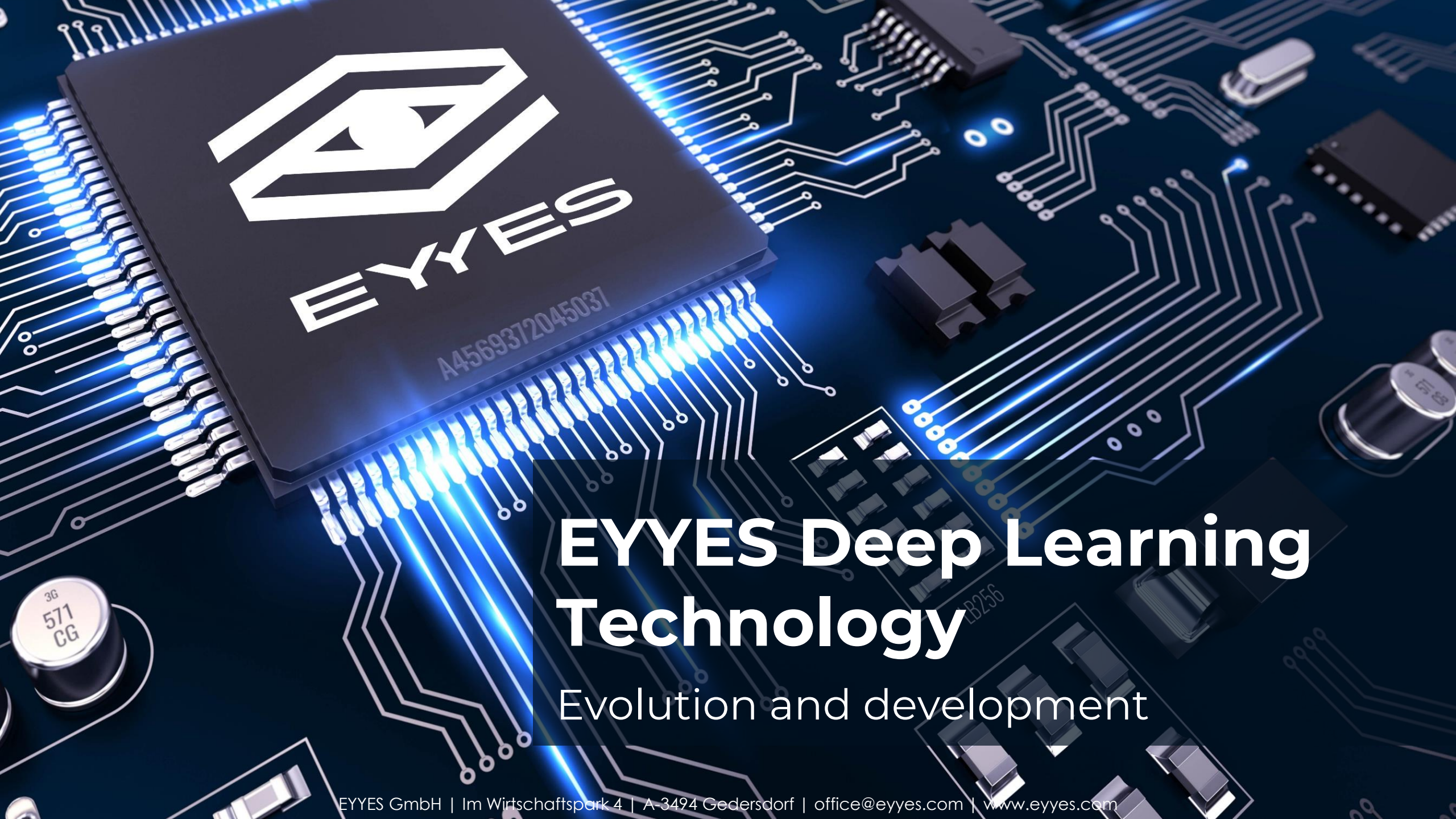


PRESENTING
Provide relevant information at the right time

DETECTION
With high-resolution sensors detect every detail

INTEGRATING
Running on EYYES hardware energy or platform independent

we make machines see



EYYES Deep Learning Technology

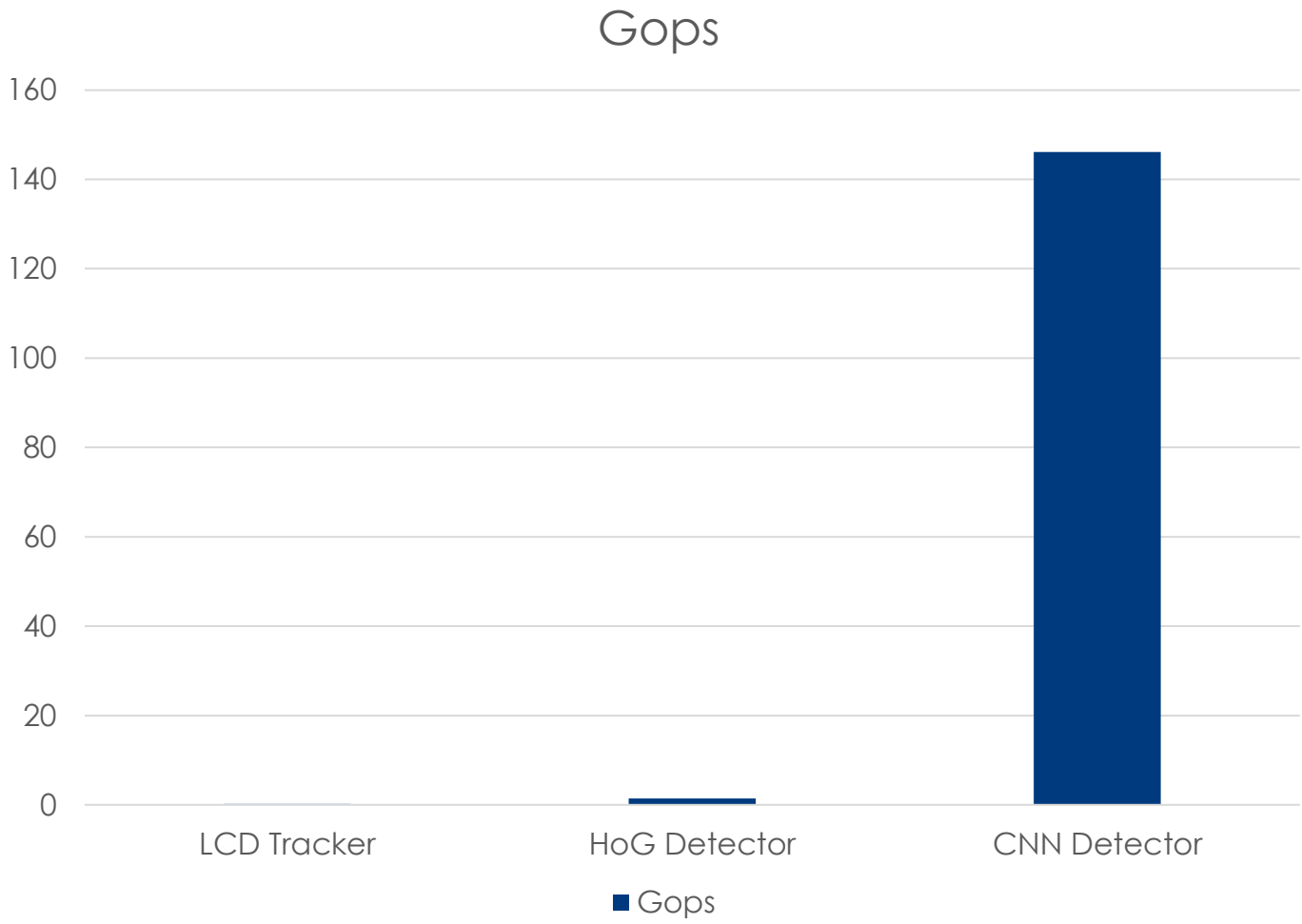
Evolution and development

EYYES GmbH | Im Wirtschaftspark 4 | A-3494 Gedersdorf | office@eyyes.com | www.eyyes.com

Initial Situation in 2016

Deep Learning needed for several project opportunities

- 2 powers of ten more calculation requirement
- On the edge of the physical limitation
- GPU require very high electrical power consumption and produce lots of waste heat
- No real alternative avaiable

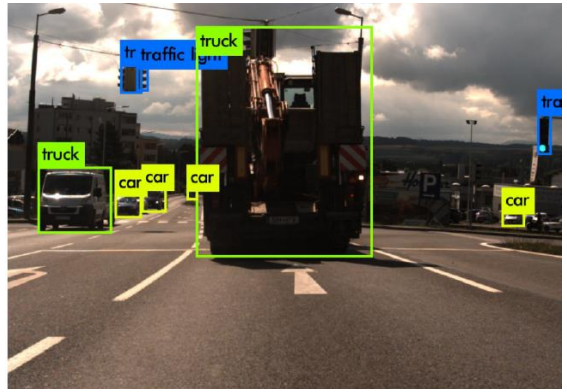


we make machines see

EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

Projects realizes with public and private funding:



- **AIRVS „Artificial Intelligence Rear View System“** together with SCCH Hagenberg
  - YOLO based network tests
  - Research on LSTM based tracking algorithms
  - Development of CNN software optimization algorithms

- **RailEye 3.0** in cooperation with TU-Dresden
  - Development of a SoM for 2 sensor realtime applikations
  - FPGA implementation of H.264 core and first deep-learning processing

EYYES we make machines see

SAB

FFG

TECHNISCHE UNIVERSITÄT DRESDEN

scch software competence center hagenberg

EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

How to solve the challenge of maximizing the performance of a CNN chip?

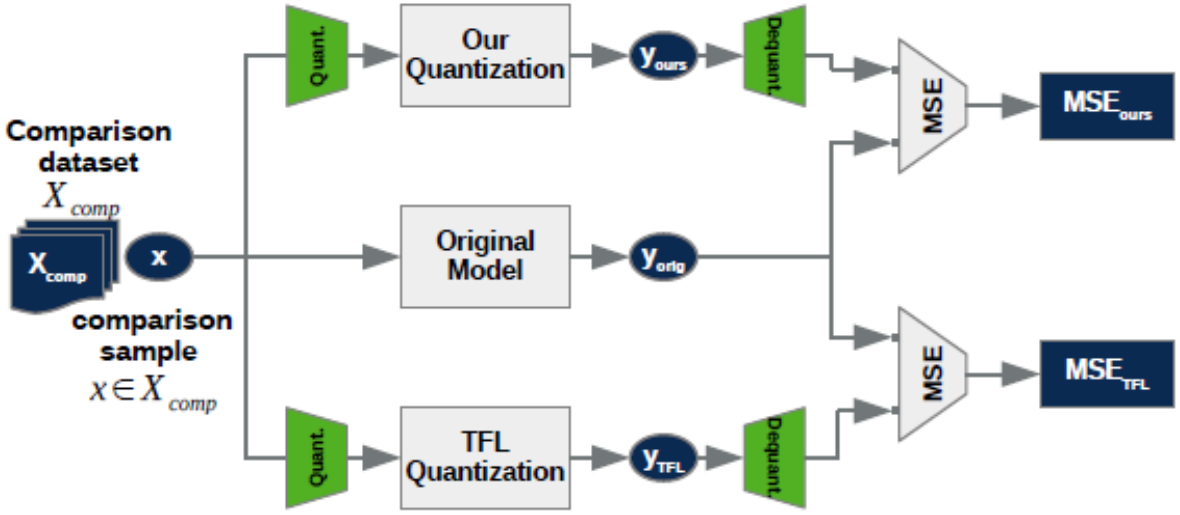
- Use quantisation to reduce the required memory bandwidth
- Decrease the required operations by second
- Improve the training algorithm
- Use explainability algorithm to monitor the functionality of the neural network
- Improve the parallel processing

EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

Challenge 1: Use quantisation to reduce the required memory bandwidth

- EYYES developed a new approach to quantize the CNN parameter (**patent pending**)
- Methode to determine the meansquare error (MSE)



TFL = Tensorflow Lite, https://www.tensorflow.org/lite/
Ours = EYYES pptimization Toolchain, https://www.eyyes.com/technology/deep-learning-optimizer/

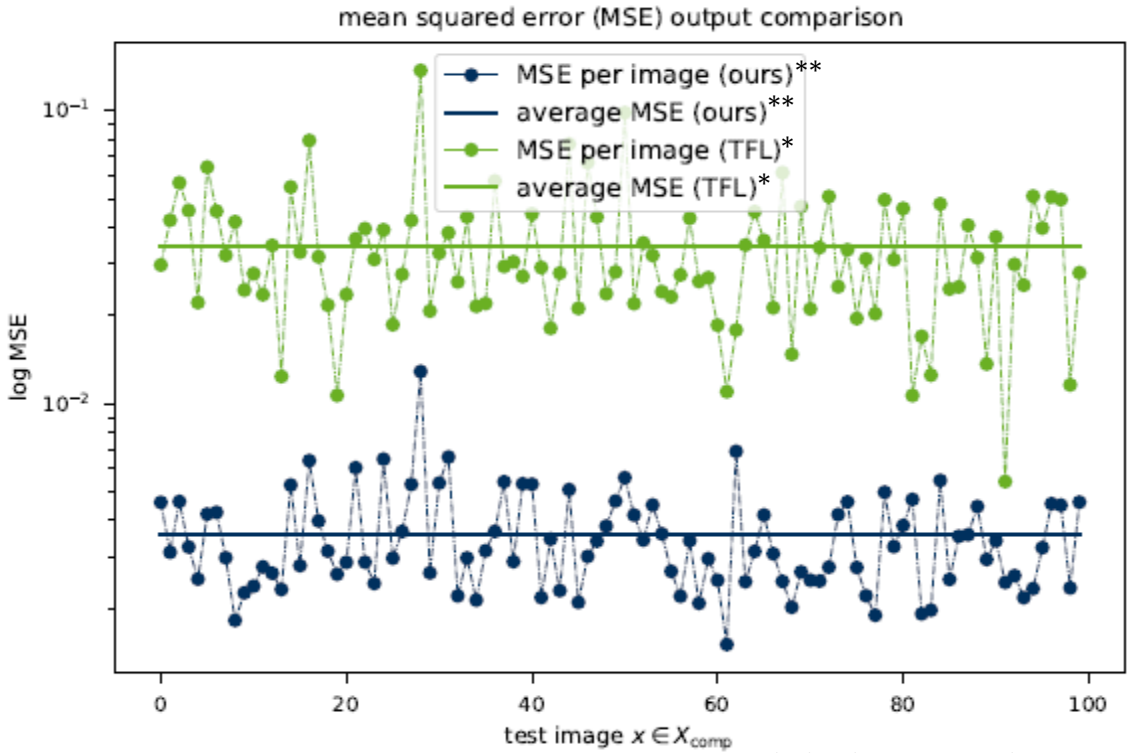
we make machines see

EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

Challenge 1: Use quantisation to reduce the required memory bandwidth

- Results comared with Tensorflow lite (TFL*)



	average	minimum	maximum	σ
MSE_{TFL}	$3.4 \cdot 10^{-2}$	$5.4 \cdot 10^{-3}$	$1.4 \cdot 10^{-1}$	$1.9 \cdot 10^{-2}$
MSE_{ours}	$3.6 \cdot 10^{-3}$	$1.5 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$	$1.5 \cdot 10^{-3}$

**Ours = EYYES optimization Toolchain, https://www.eyyes.com/
* TFL = Tensorflow Lite, https://www.tensorflow.org/lite/

we make machines see

EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

Challenge 2: Decrease the required operations by second

- Reduce the required operations using
  - Pruning
  - Cutting
  - Specific additional reductions

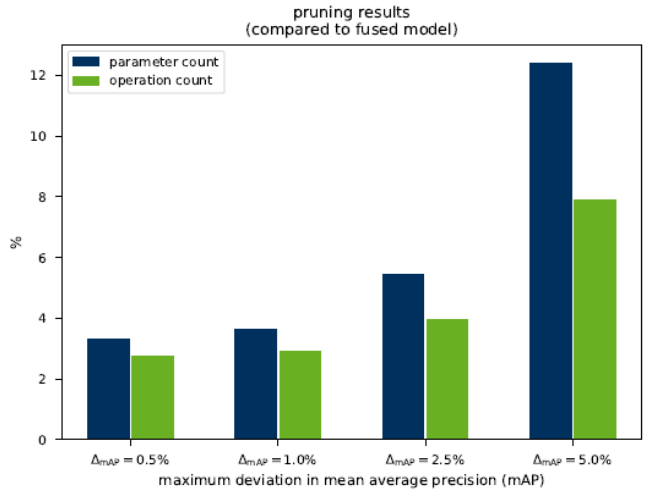
mAP deviation: Δ_{mAP}	*Parameter reduction:* $R_{N_{\text{param}}}$	*Operation reduction:* $R_{N_{\text{ops}}}$
0.5%	3.3%	2.7%
1.0%	3.6%	2.9%
2.5%	5.4%	4.0%
5.0%	12.4%	7.9%

EYYES Deep-Learning approach

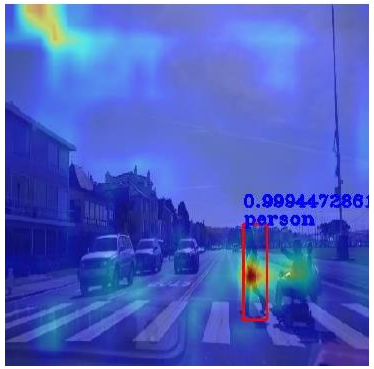
Soft- and Hardware R&D Projects realized

Challenge 3: Improve the training algorithm

- EYYES developed unique training mechanism
  - Autoannotation
  - Quality measurement (MaP, IoU, …)
  - Simulation of the network
  - Perturbation methods to challenge the DNN
  - Extend the variety of objects and noise using „Prototypes“ and GANs
  - Explainability due to stepwise analysis methods



a

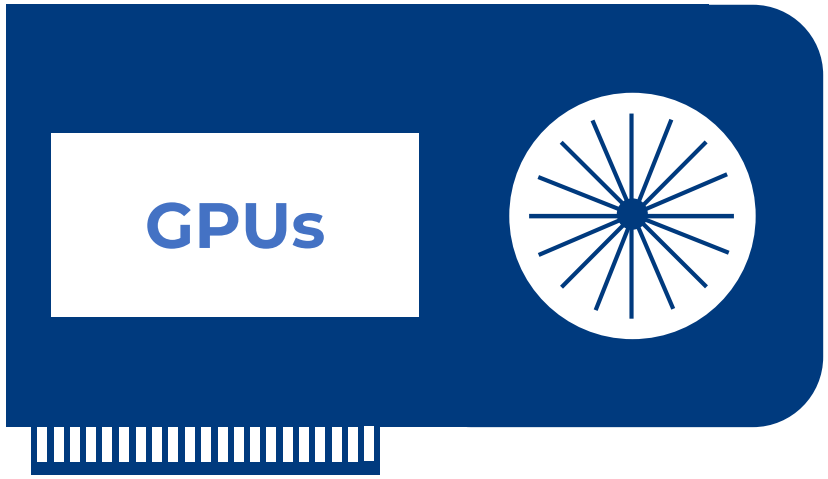


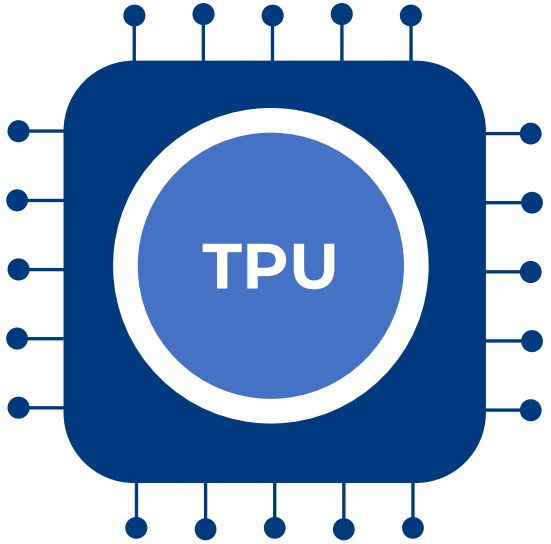
we make machines see

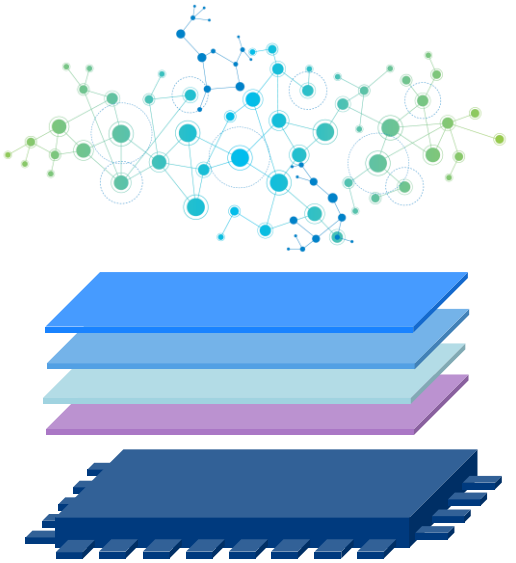
EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

Challenge 4: Improve the parallel processing







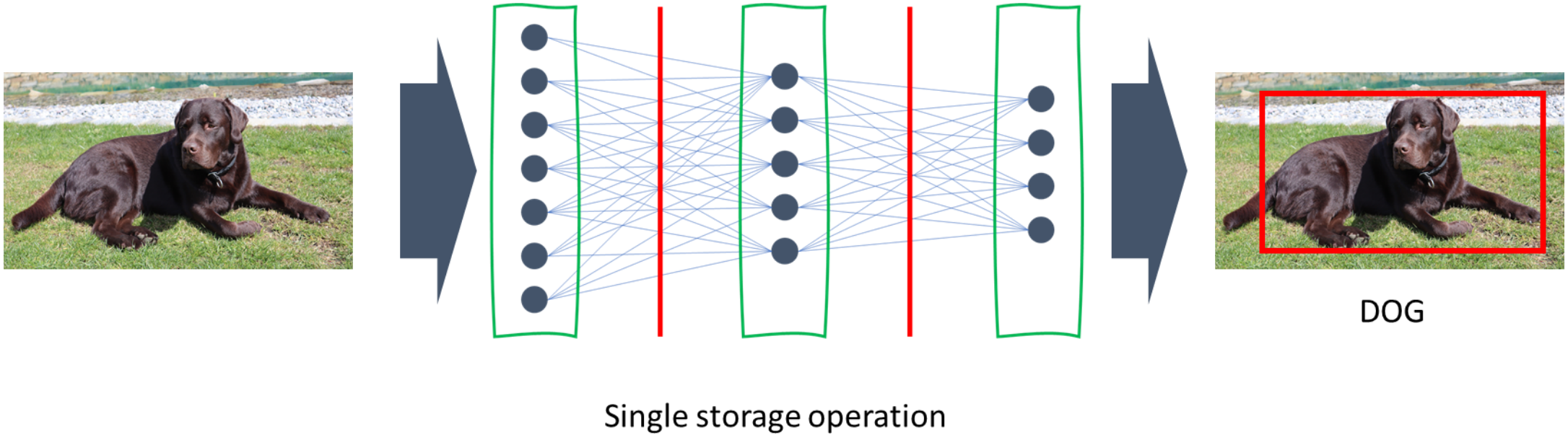
LPU

we make machines see

EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

Challenge 4: Improve the parallel processing



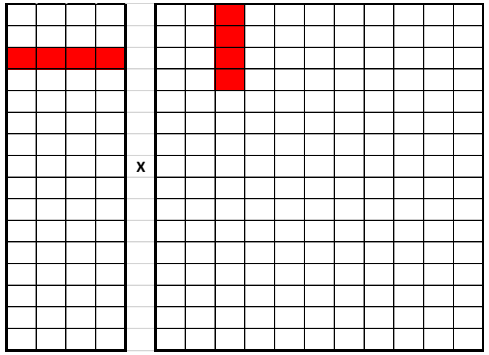
we make machines see

EYYES Deep-Learning approach

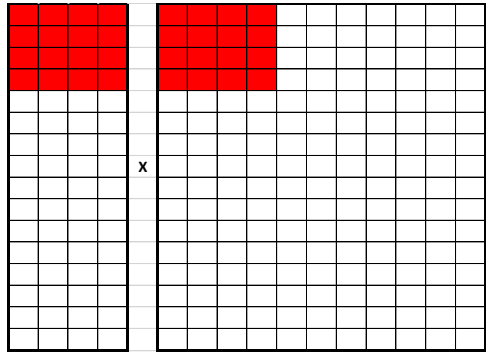
Soft- and Hardware R&D Projects realized

Challenge 4: Improve the parallel processing

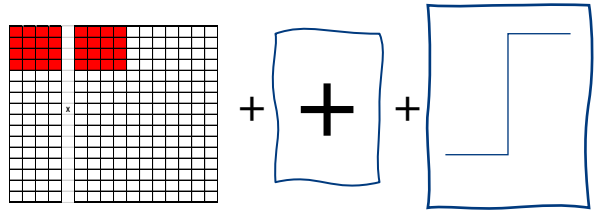
GPU
(Vector operations)



TPU
(Tensor operations)



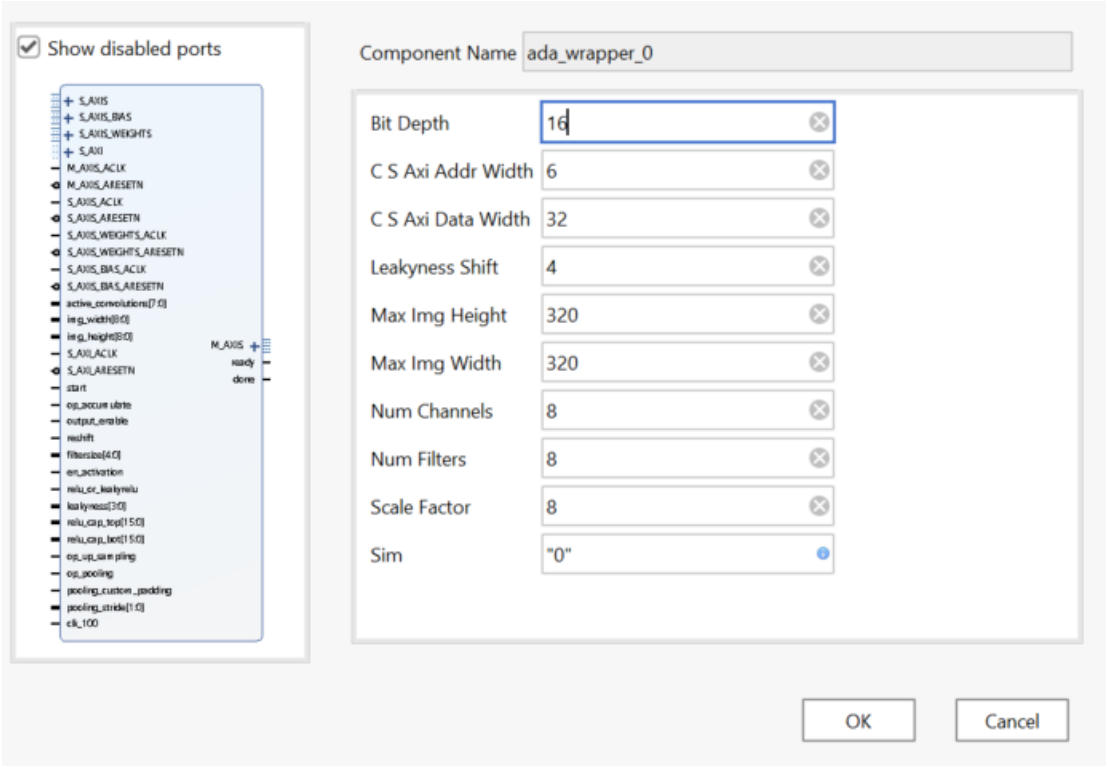
LPU
(Layer operations)



EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

Challenge 4: Improve the parallel processing



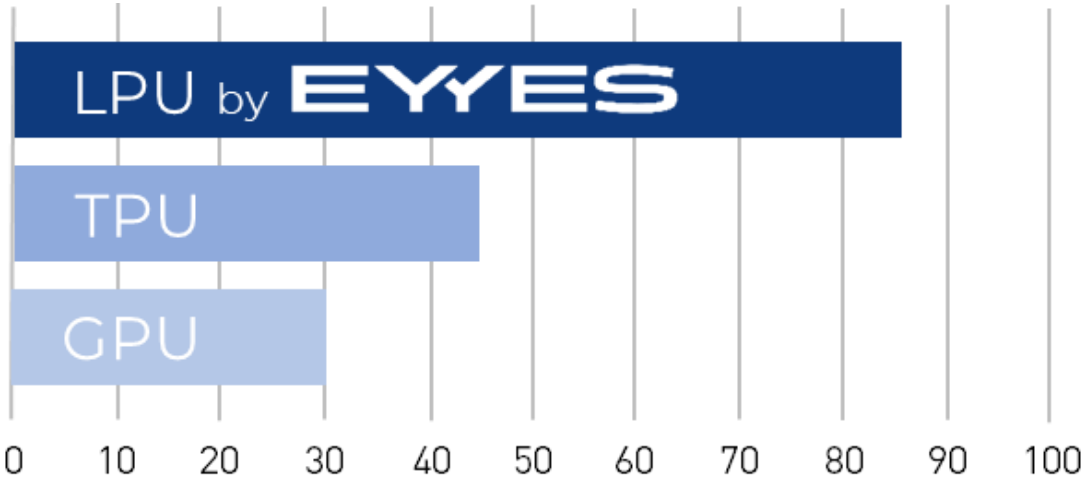
we make machines see

EYYES Deep-Learning approach

Soft- and Hardware R&D Projects realized

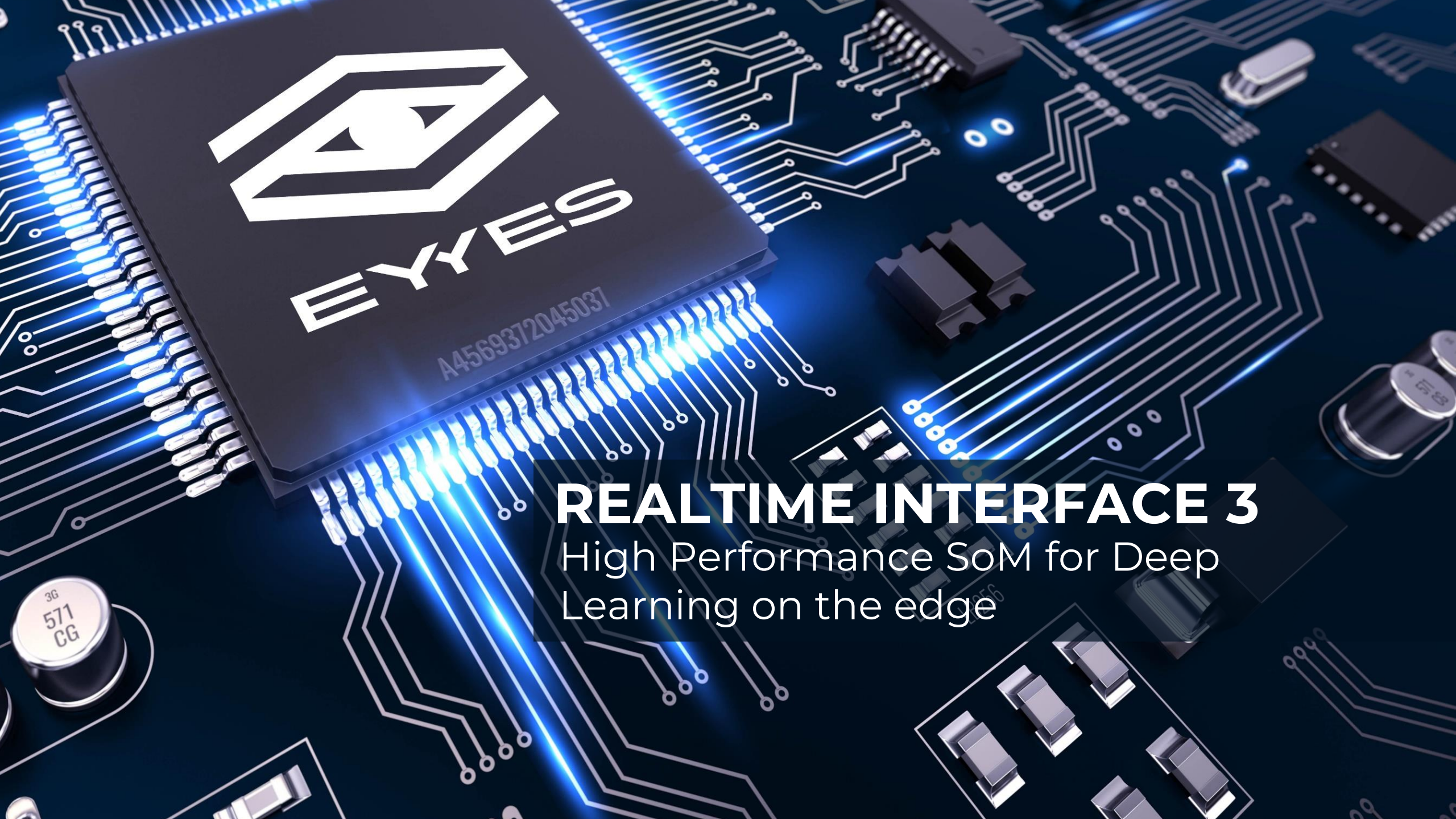
Challenge 4: Improve the parallel processing

- Maximum parallelism
- Gerneralized processing unit
  - Kernel W 1-16, H 1-16
  - Strides W 1-2, H 1-2
  - Padding 0
  - Maxpooling
  - Fully connected
  - Input Size arbitrary
  - Convolution and depthwise convolution
  - Up to 32 Cores
  - > 10.000 operations per clock



LPU Terra Operations per Second compared between the LPU, TPU and GPU using similar frequencies

we make machines see

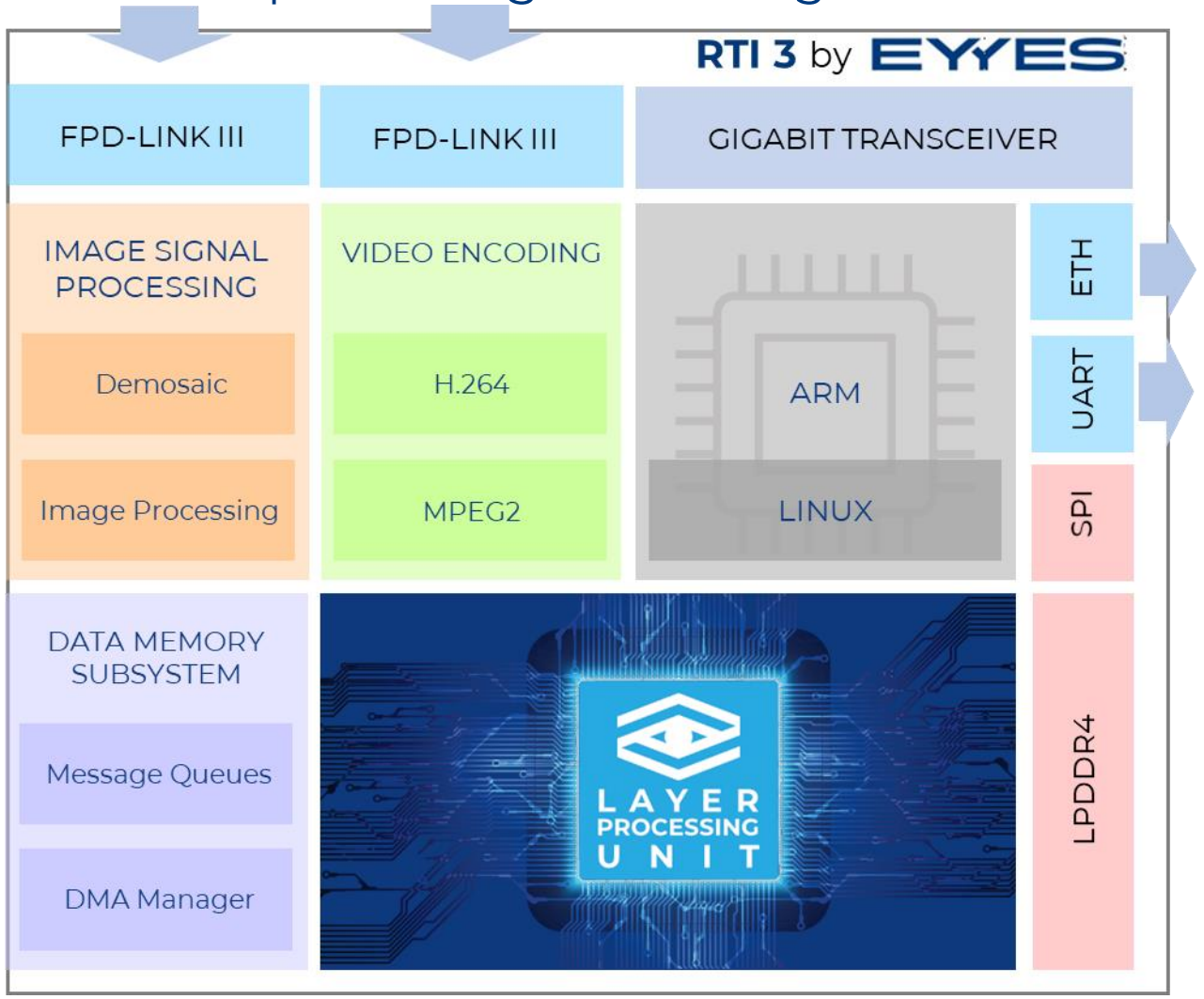


REALTIME INTERFACE 3

High Performance SoM for Deep Learning on the edge

REALTIME INTERFACE - RTI 3

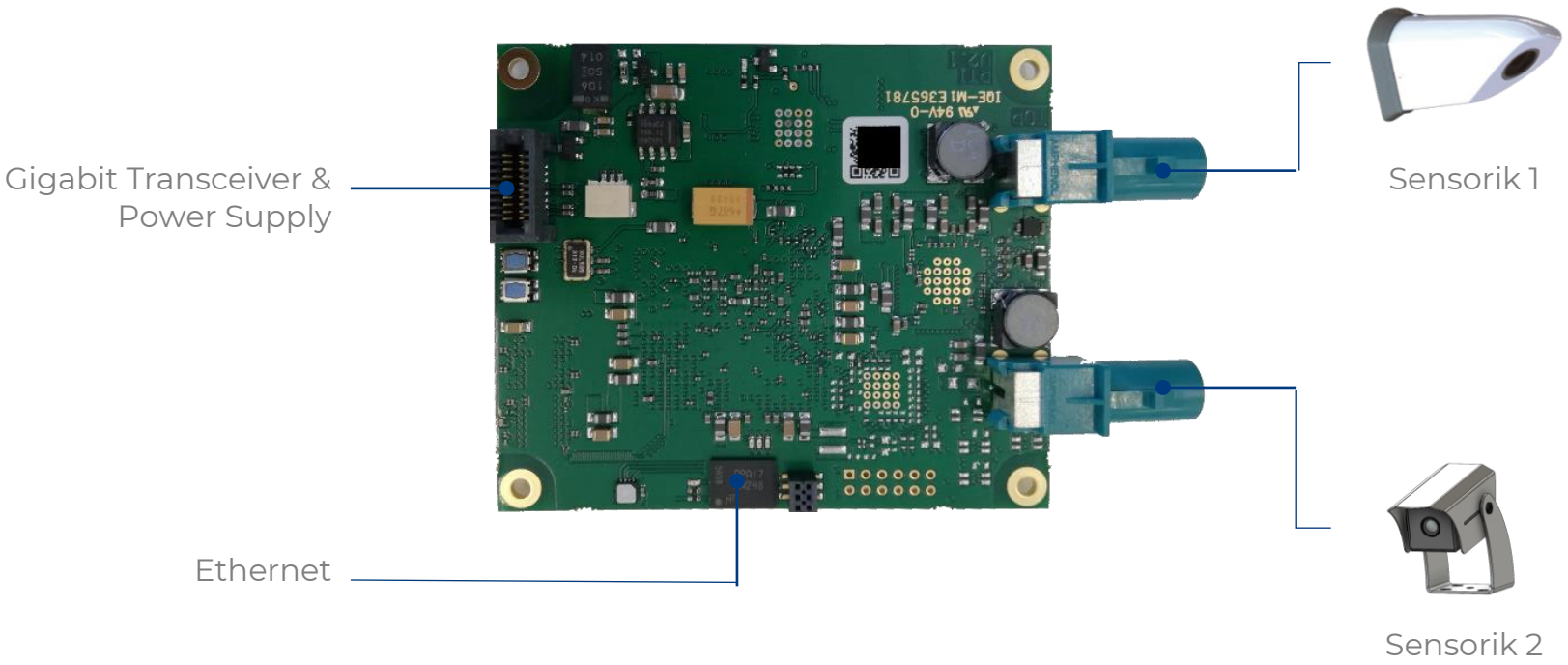
High Performance SoM for Deep Learning on the edge



we make machines see

REALTIME INTERFACE - RTI 3

High Performance SoM for Deep Learning on the edge

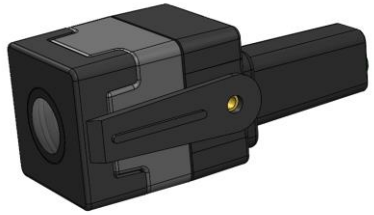


we make machines see

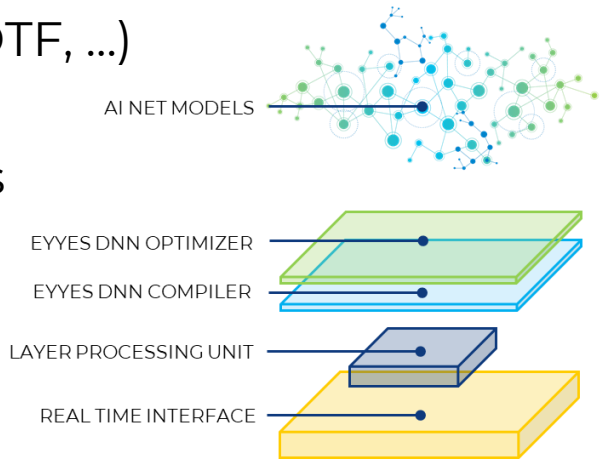
REALTIME INTERFACE - RTI 3

High Performance SoM for Deep Learning on the edge

The perfect Deep-Learning platform:



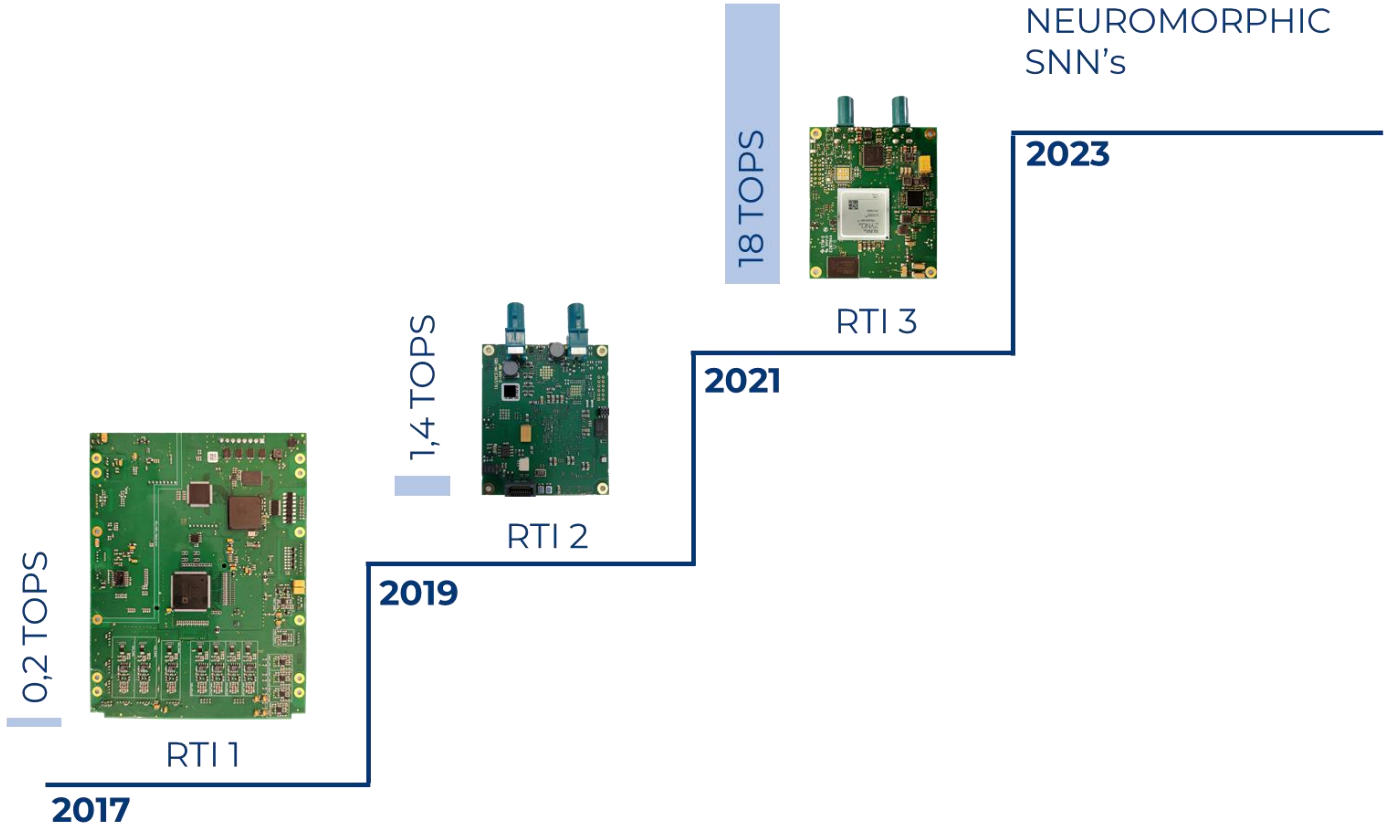
- Plug&Play device together with the EYYES camera sensors
- Power Supply via Power over Ethernet or direct power supply (low power)
- Process and control up to two independent Camera sensors via FDP LINK III
- Process up to two different digital H.264 videostreams
- Receive the object list directly with open standard protocol (ROS, ADTF, ...)
- Easy to configure using Webinterface (easy to use)
- Process in realtime the sensor data with deep-learning with 20 TOPs
  - Preinstalled EYYESNET with 7/21 object classes
  - Specialization and replacement of the DNN via Update



we make machines see

EYYES Technology Evolution

FPGA Driven Development and Outlook



Evolution from an RTI1 to RTI3 and Outlook

REALTIME INTERFACE - RTI 3

High Performance SoM for Deep Learning on the edge

Examplevideo from Testdrive in Vienna

we make machines see