Institute of Microelectronic Systems

Leibniz Universität Hannover

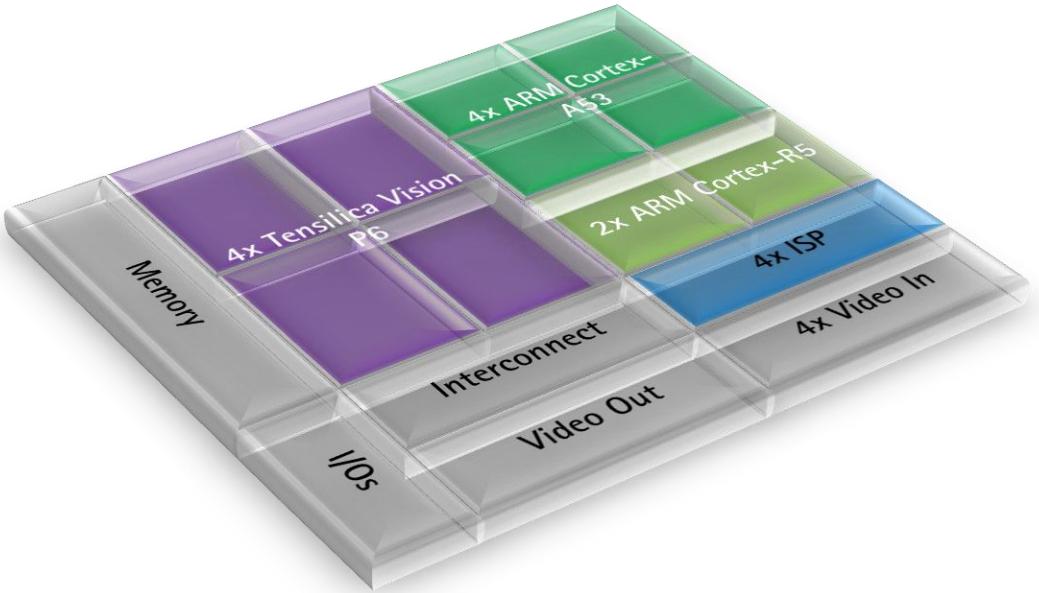
High Performance, Energy Efficient Computer Vision for ADAS on Tensilica Vision P6

Tensilica Day 2017

Nicolai Behmann, M.Sc.
Prof. Dr.-Ing. Holger Blume

Dream Chip *Software Defined Image Processing* (SDIP) SoC

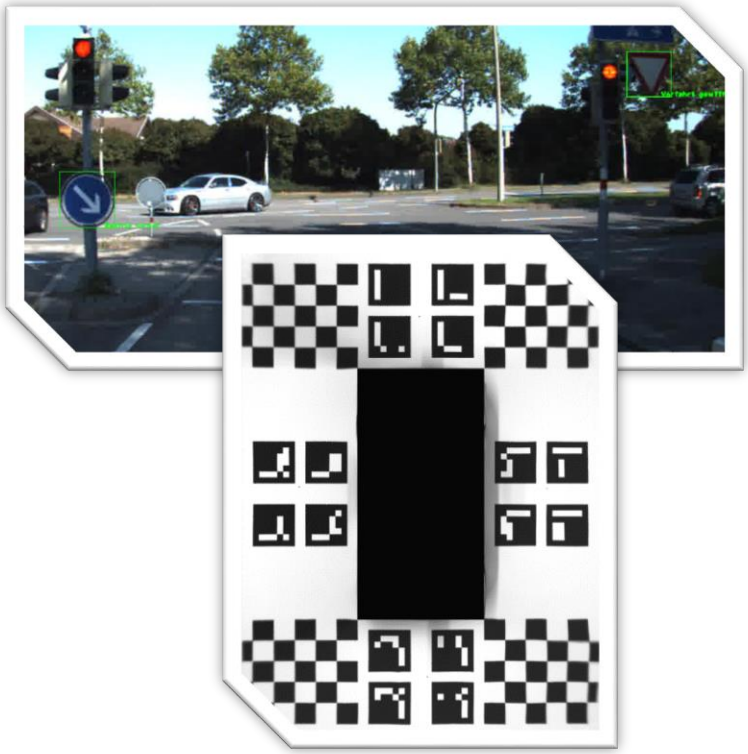
- *ARM Cortex-R5* Lock Step
  - ADAS safety processor
- 4x *ARM Cortex-A53*
  - General purpose processor
  - NEON SIMD ext.
- 4x *Tensilica Vision P6*
  - Computer Vision Acc.
- 4x Video in with *ISP* pipeline
- 2x LPDDR4 memory



Outline – Accelerated ADAS Applications

- Tensilica Vision P6 architecture

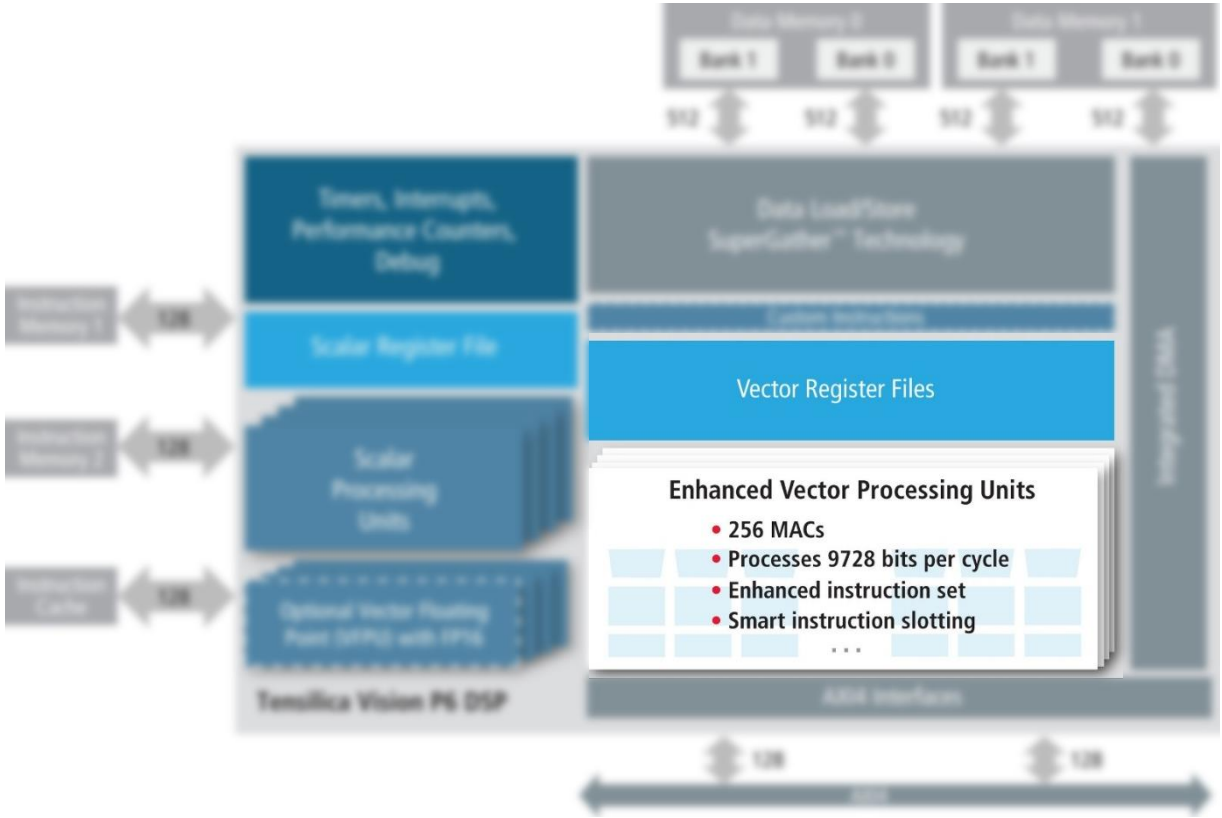
- Case studies
  - *Convolutional Neural Network (CNN)*: Traffic Sign Recognition Benchmark
  - *360° Surround View*



- Implementation guidelines
  - HW-SW Co-Design
  - SIMD parallelization
  - DMA usage
  - XI library

Tensilica Vision P6 architecture

- 512b/1536b SIMD (64x8b, 32x16b, 16x32b)
- 1-5 issue slot VLIW
- Up to 4x ALU, 1 MUL/MAC, 1 Select, 2 loads or 1 load + 1 store per cycle



Tensilica Vision P6 architecture

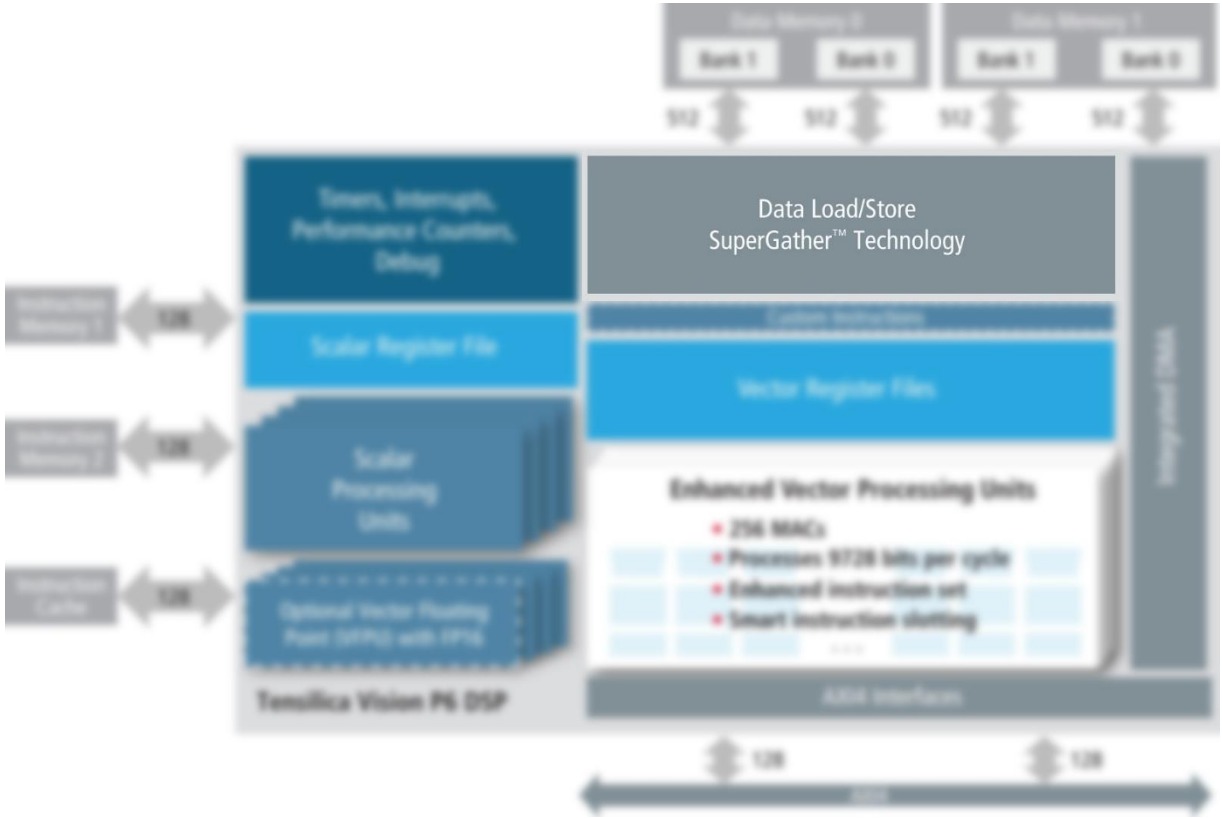
- 2 separate local memorys
- 2 banks each (interleaved)
- 8 sub-banks per bank
- SDIP: 2x 256kB

Tensilica Vision P6 architecture

- Hiding system memory latency
- 1D + 2D transfers
- System to local memory, local to system memory, local to local memory

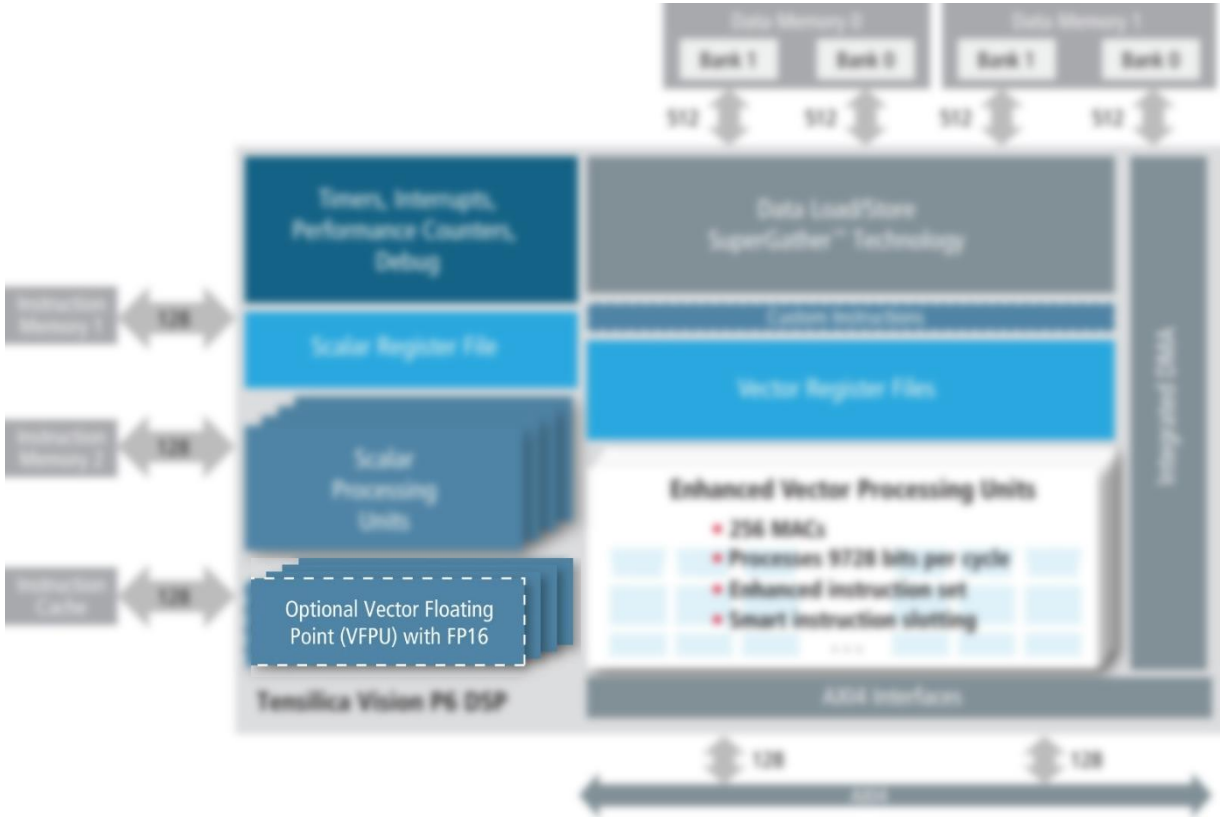
Tensilica Vision P6 architecture

- Read non-contiguous locations from memory
- Up to 32 elements / cycle
- non-blocking



Tensilica Vision P6 architecture

- 16-way 32b vector FP
- 32-way 16b vector FP
- Supports autovectorization, intrinsics

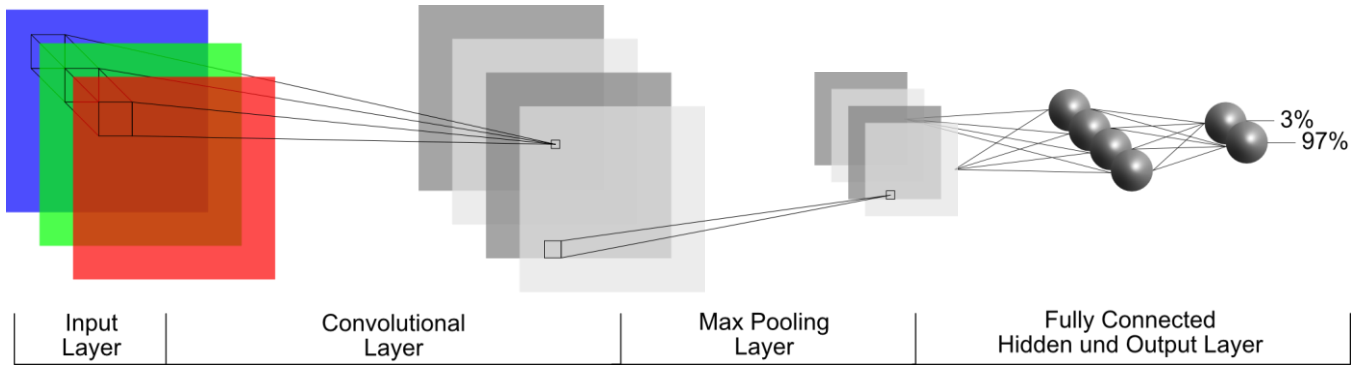


Case study: Convolutional Neural Network (CNN)

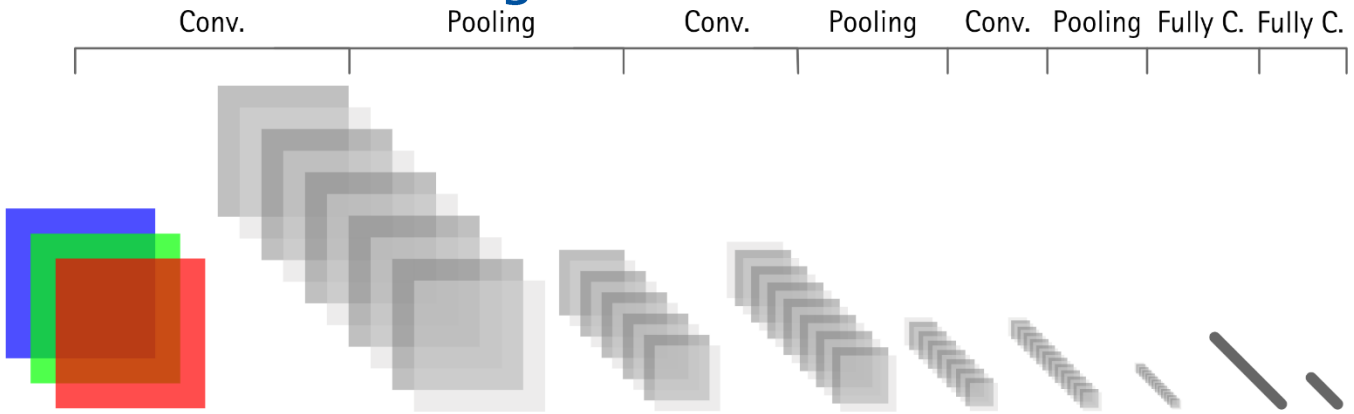
- Convolutional Layer: local feature extraction
  - 2D-Convolution of input feature map volume with pretrained filter kernels
  - Subsequent pixel-wise transformation with non-linear activation function

- Max Pooling Layer: non-linear information reduction
  - Maximum value from local 2x2 neighborhood

- Fully Connected Layer: final classifier
  - Connection of each neuron with each neuron on previous layer
  - Class propability represented in output vector



CNN architecture Co-Design

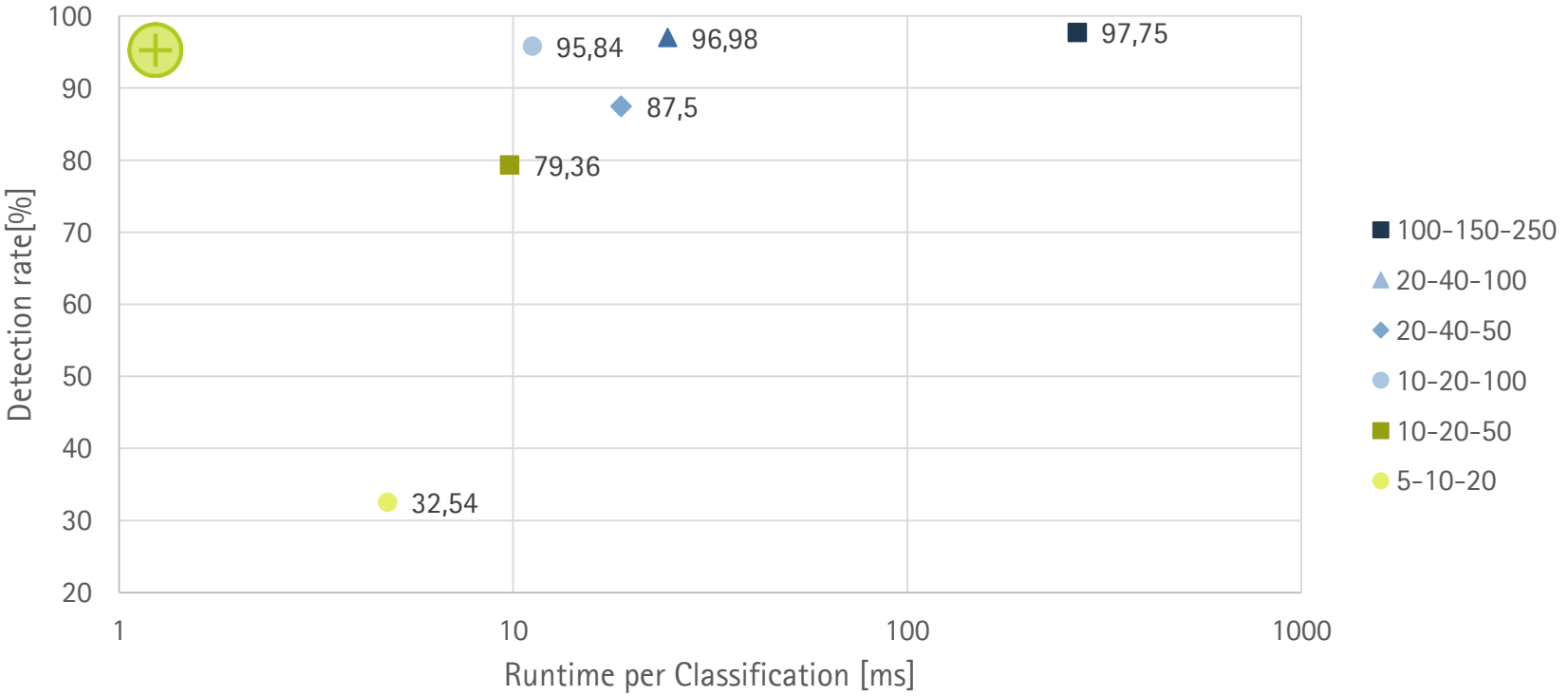


Neuronen	48x48	42x42	21x21	18x18	9x9	6x6	3x3	300	43
Filter-Kern		7x7	2x2	4x4	2x2	4x4	2x2		
Feature Maps nach Ciresan[1]		100	100	150	150	250	250		
20-40-100		20	20	40	40	100	100		
20-40-50		20	20	40	40	50	50		
10-20-100		10	10	20	20	100	100		
10-20-50		10	10	20	20	50	50		
5-10-20		5	5	10	10	20	20		

- Problem: Convolutional Layer responsible for 99% of computing load, memory bandwidth limited
- Training of different CNN architectures with less feature maps

German Traffic Sign Recognition Benchmark[2] results

- Training with more than 40.000 traffic signs of 43 classes, randomly transformed
- Runtime for Intel Core i5 @ 2.6 GHz



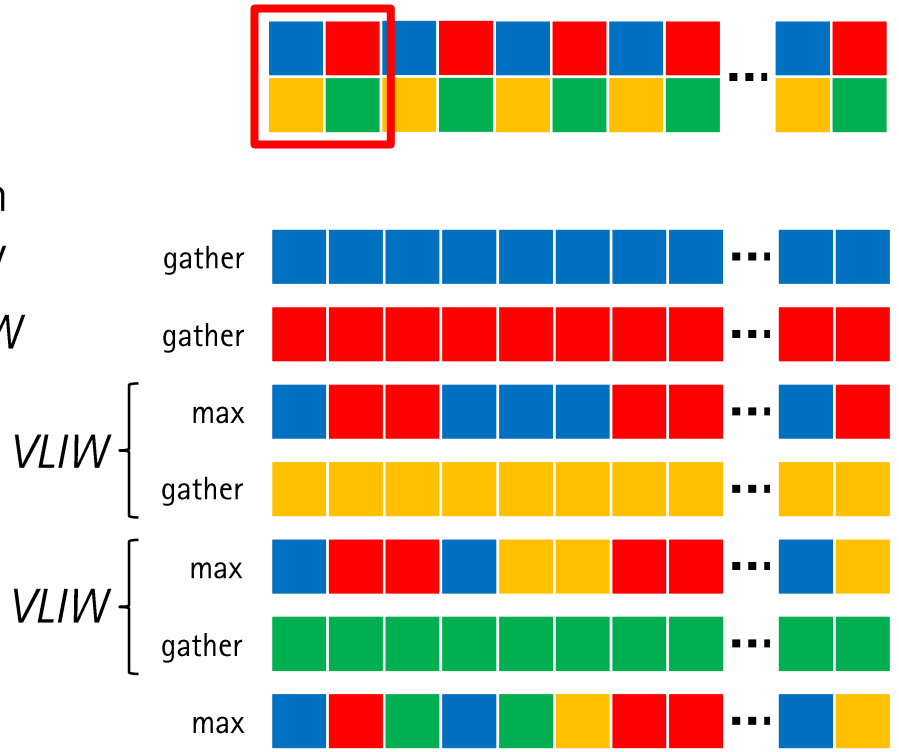
➢ Acceleration of CNN with Tensilica Vision P6 DSP

[1] Houben-IJCNN-2013

CNN: Tensilica Vision P6 implementation

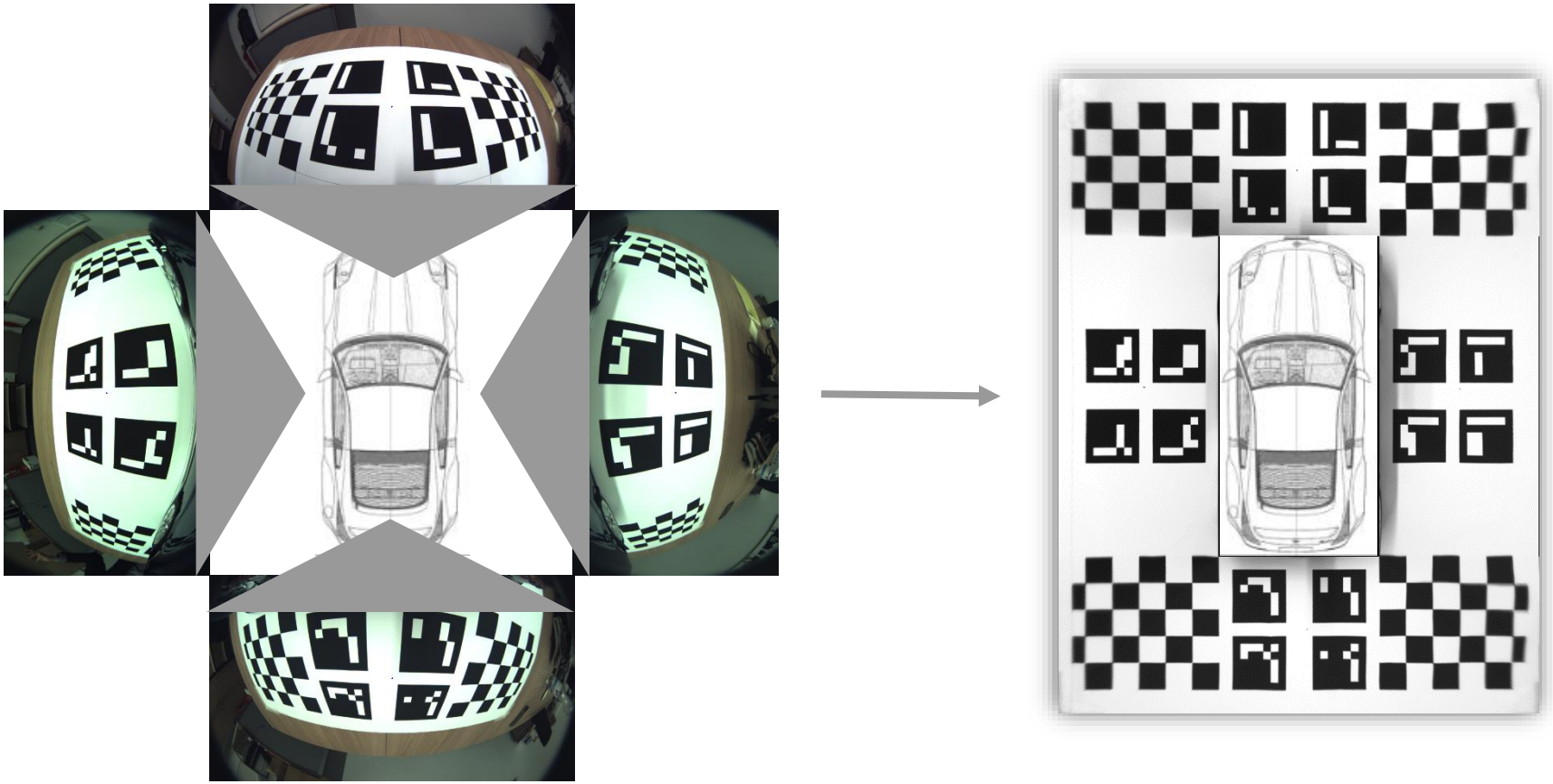
1. Floating-point to Fixed-point conversion (small increase in detection rate noticeable)
2. Local DRAM for fast memory accesses
3. Concurrent preloading of local memory with DMA engine (double buffering)
4. SIMD vectorization of CNN layers: example Max-Pooling of 2x2 local neighborhood

- Performance results
  - 95,7% detection rate
  - Speedup of 64 through optimization
  - VP6: 330 CNNs / s @ 750 MHz, *4 W*
  - Core-i5: 83 CNNs/s @ 2,6 GHz, *45 W*



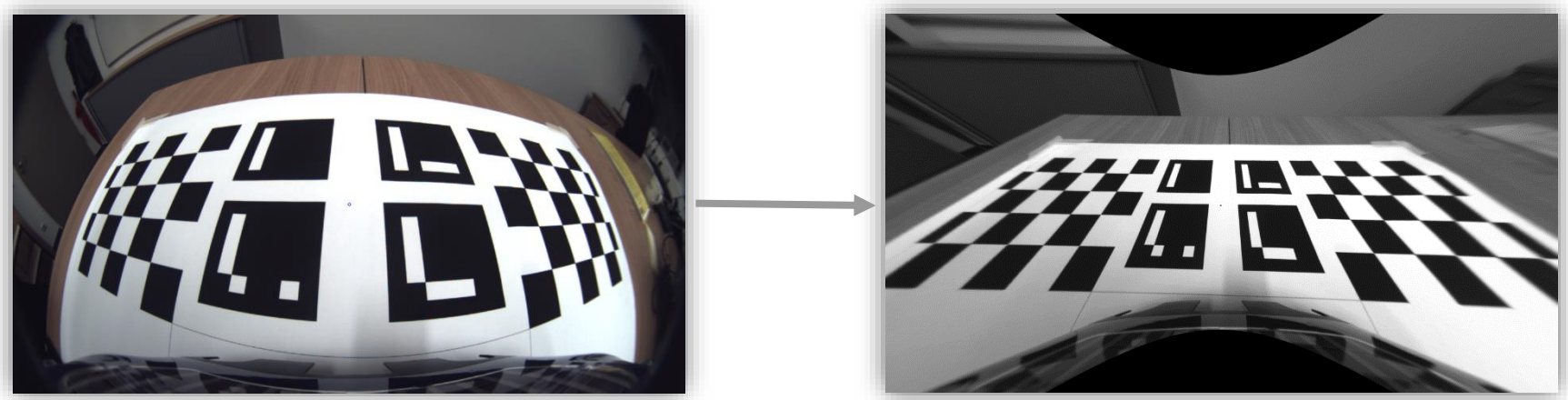
Case study: 360° Surround View

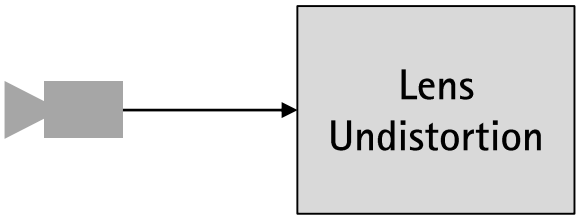
- Goal: Synthetic bird view perspective (e.g. ADAS for parking maneuvers)
- Input: 4x fisheye camera around the car



360° Surround View: Pipeline 1

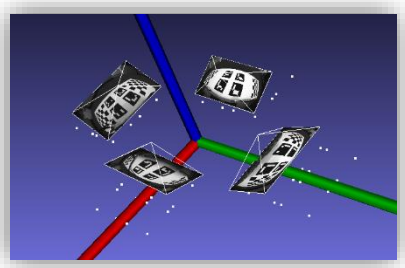
- Lens Undistortion: Fisheye model (polynomial approximation)
  - $\theta_d = \theta(1 + k_1\theta^2 + k_2\theta^4 + k_3\theta^6 + k_4\theta^8), \theta = \text{atan}(r)$

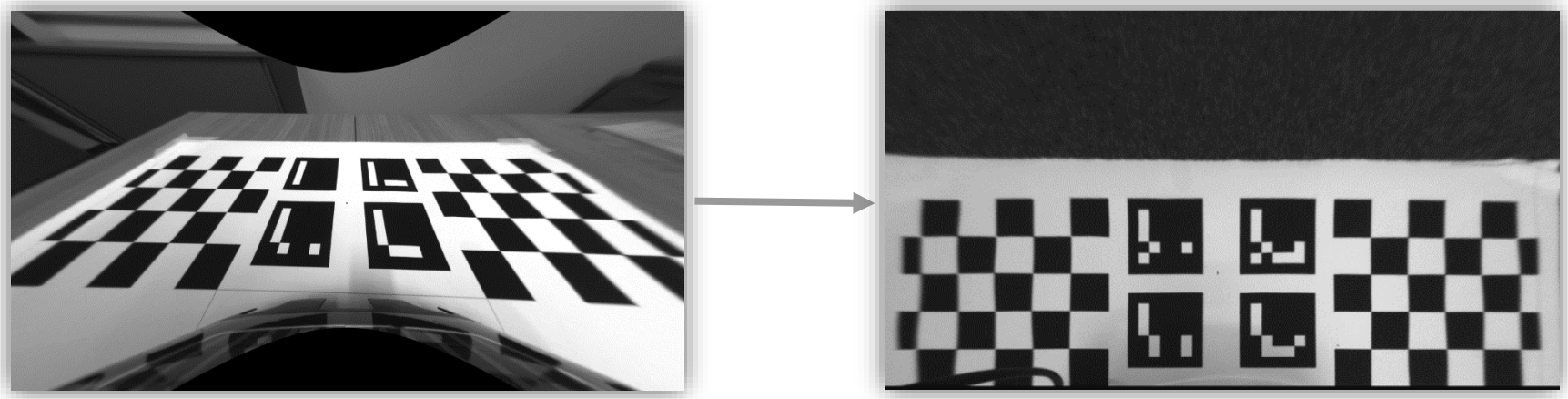


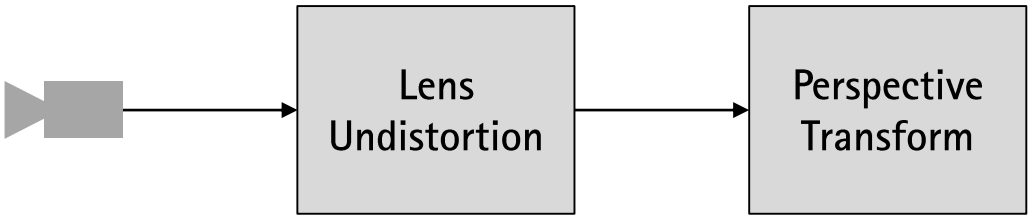


360° Surround View: Pipeline 2

- Perspective Transformation: affine remapping to „top view“
  - $\underline{x}_{top} = \begin{bmatrix} \underline{\underline{R}} & \underline{T} \\ 0 & 1 \end{bmatrix} \cdot \underline{x}_{undistorted}$

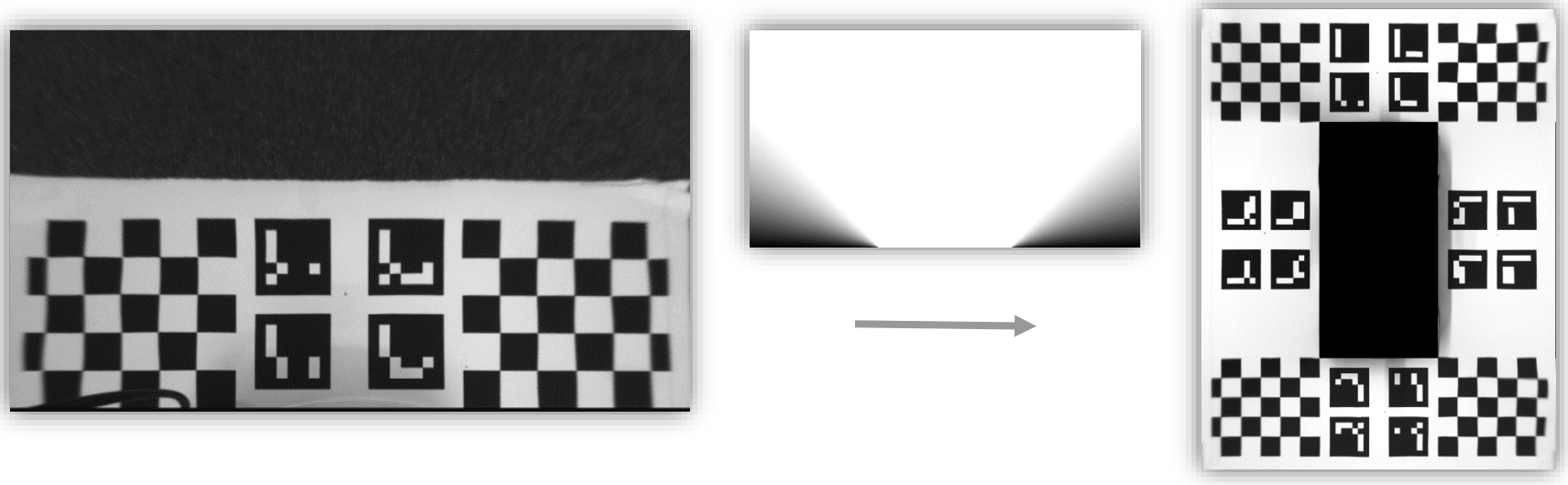






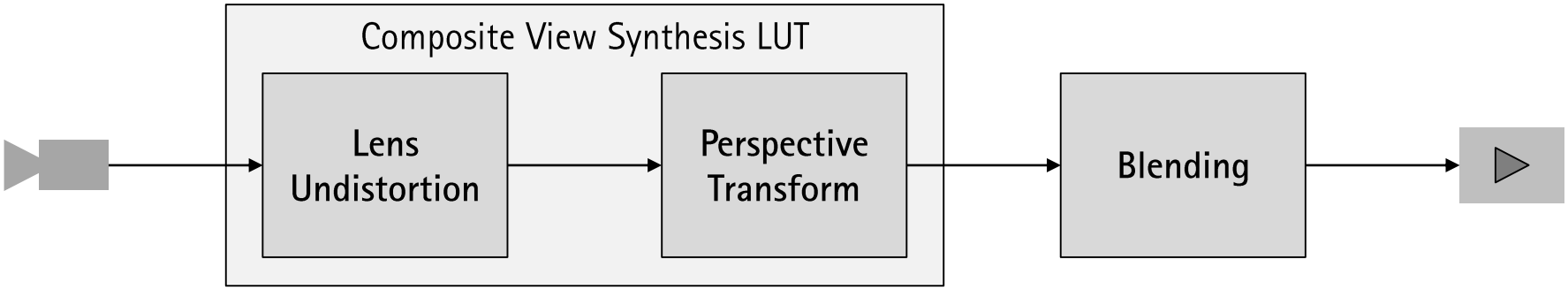
360° Surround View: Pipeline 3

- Pixelwise multiplication of topview image with mask
- Overlay of all blended images



360° Surround View: Look-Up table approach

- Assumption: Static remapping functions and blending masks
  - Combination of remapping steps into one



360° Surround View: Composite View LUT Implementation

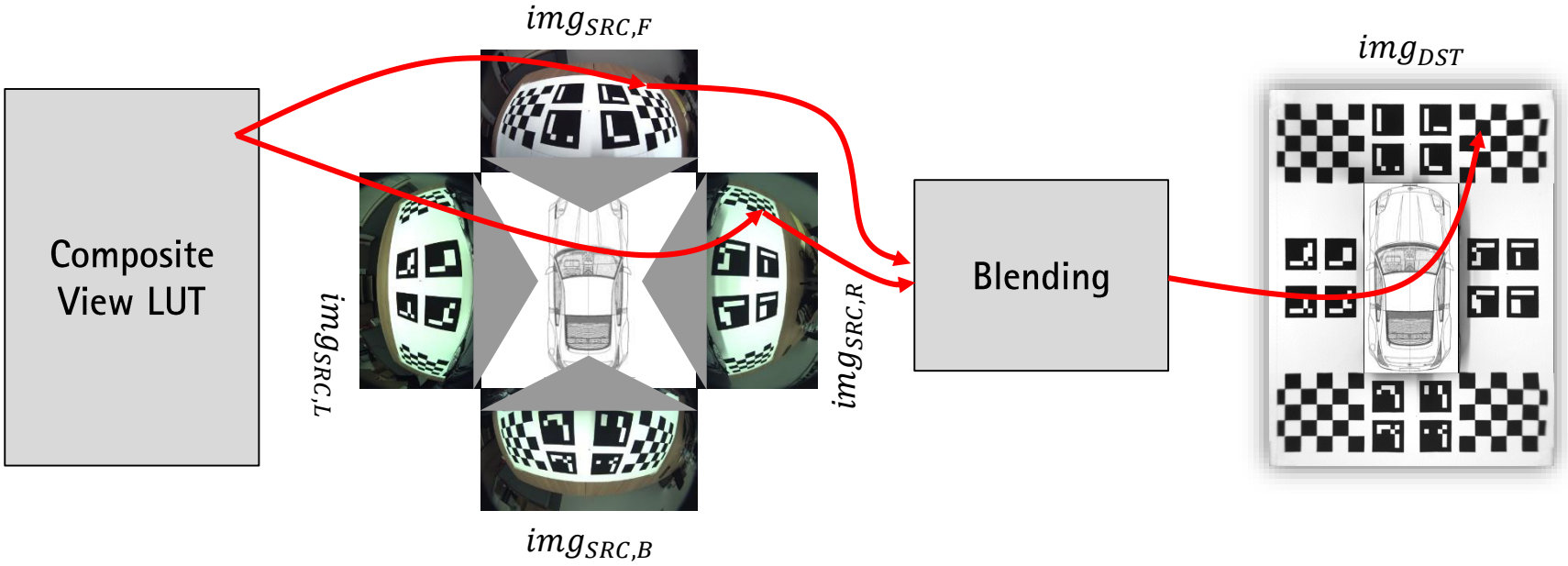
For each pixel $(I_{DST}(x,y))$ in img_{DST}

$$I_{SRC1} = I_{SRC,LUT_{CAM1}(x,y)}(LUT_{POS1X}(x,y), LUT_{POS1Y}(x,y))$$
$$I_{SRC2} = I_{SRC,LUT_{CAM2}(x,y)}(LUT_{POS2X}(x,y), LUT_{POS2Y}(x,y))$$

} remap

$$img_{DST}(x,y) = \alpha(x,y) I_{SRC1} + (1-\alpha(x,y)) I_{SRC2}$$

} blending



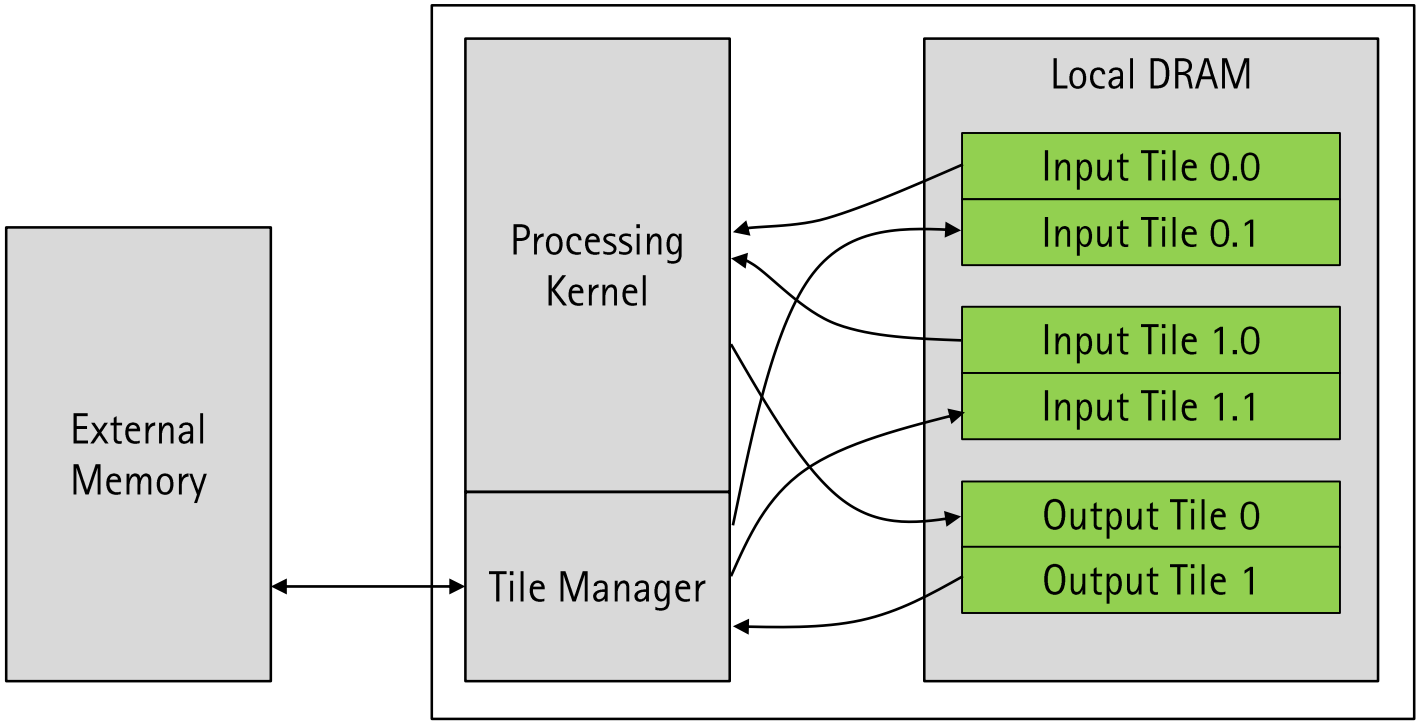
Tensilica Vision – XiLibrary

- Variety of Kernels for computer vision and video processing
- Kernels optimized for Vision Processor family architecture
- Follows OpenCV style (C-functions)
- Supports / uses different fixed-point formats
- Working of image Tiles (small image patches)
- Performance Estimation Metrics available

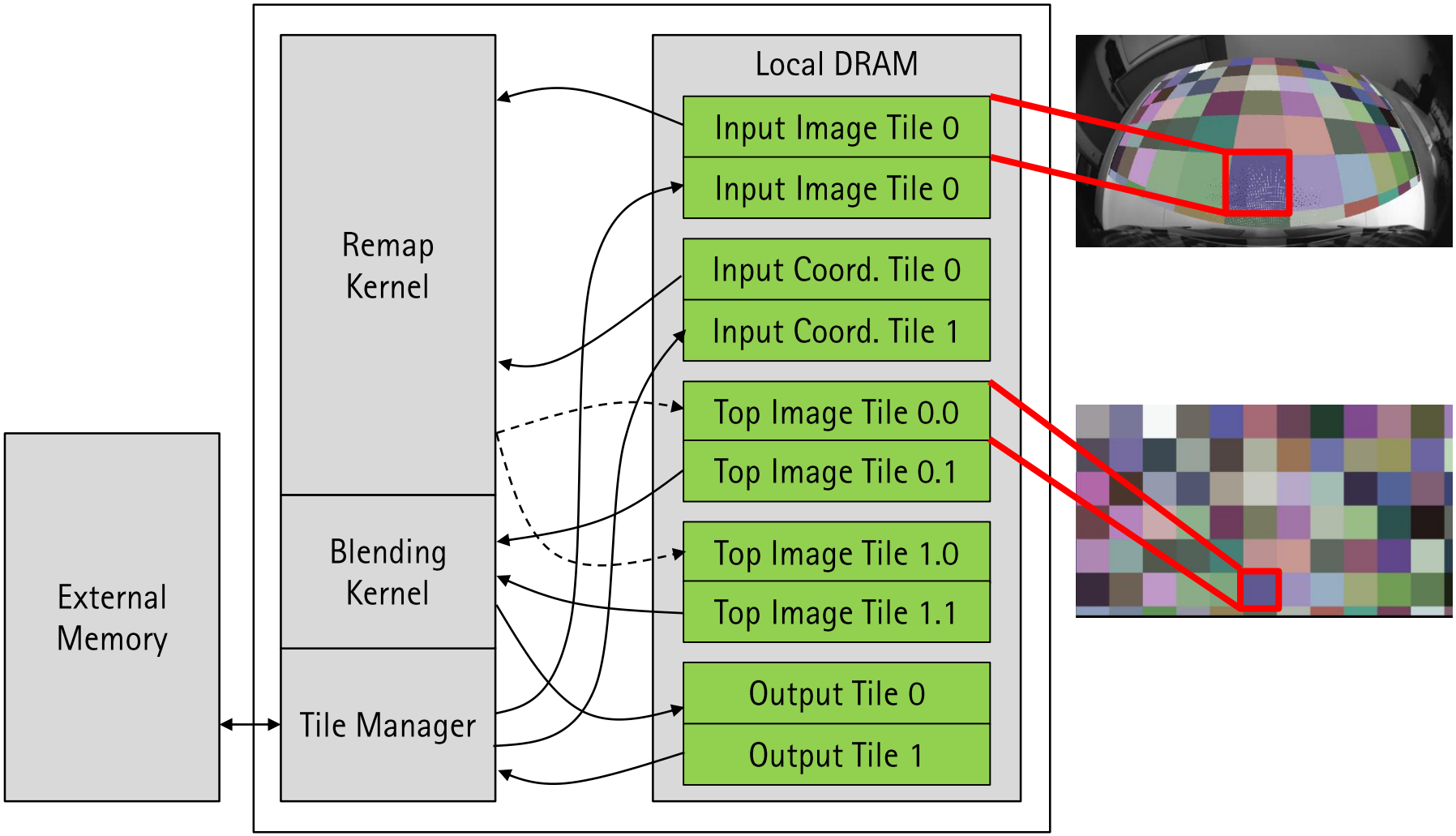
- Unary element-wise operations: abs, not, clip, neg, ...
- Binary element-wise operations: absdiff, add, addWeighted, and, ...
- Convolutions: Box, Canny, Sobel, Dilation/Erosion, Gaussian, Median, Pyramid, ...
- Geometric Transformations: flip, transpose, *warp (bilinear and nearest neighbor interp.)*
- Further functions: connected components, histogram, hough transform, integral, ...
- Feature detection, matching: Harris, Shi Thomasi, FAST, BRIEF, knn search
- Motion analysis: block matching based optical flow

Tensilica Vision – Tile Manager

- Local DRAM holds image tiles and local buffers
- Vision Processor runs processing Kernels on image Tiles
- Double Buffer is used to allow processor and DMA work in parallel
- Tile Manager keeps track of buffers in external und local DRAM and schedules DMA



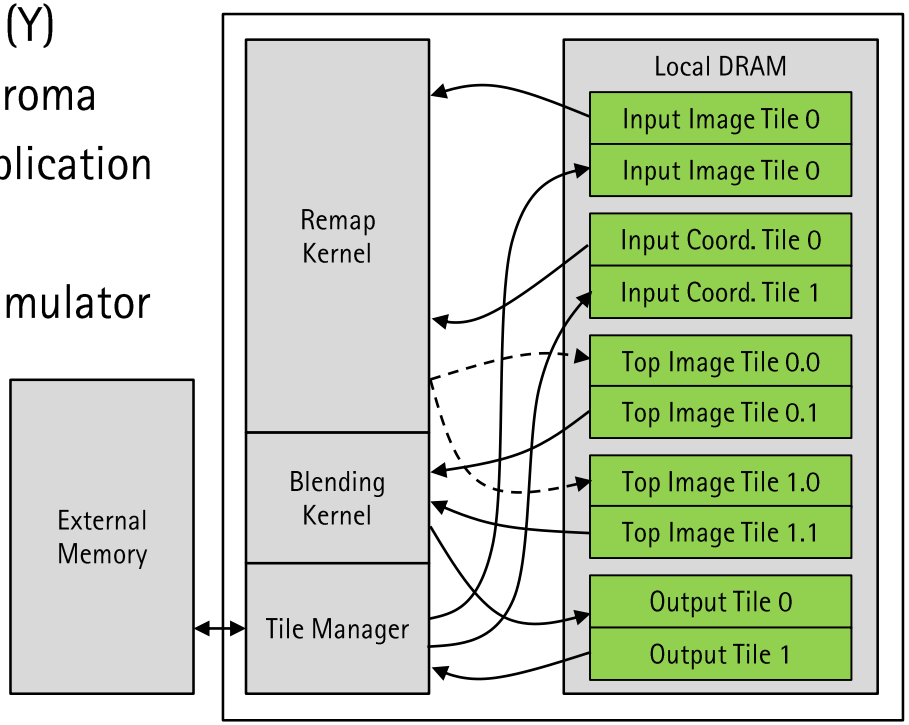
360° Surround View: Tensilica Vision P6 Implementation



360° Surround View: Tensilica Performance Results

- Remap Kernel (semi-planar YUV4:2:2 input)
  - Bilinear Interpolation for Luminance (Y)
  - Nearst Neighbor interpolation for Chroma
- Blending Kernel: element-wise Tile multiplication

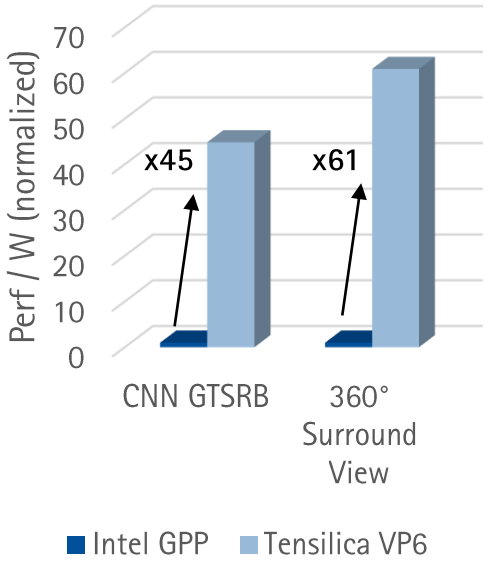
- Simulation with Xtensa Instruction Set Simulator
- Memory Latency Modelling
  - 100 cyc. 1st read, 1 further
  - 50 cyc. 1st write, 1 further



- 360° Surround View performance
  - ✓ ~10 cyc. per Pixel
  - ✓ depending on pixel position in sub-bank
  - ✓ 24,8 Mio. Cycles @ 1920x1620
  - ✓ 30 fps @ 750 MHz, *~4 W*
  - ❖ Comp. Intel Core-i7: 8 fps @ 4x3,4 GHz, *~65 W*

Conclusion – High performance energy-efficient Computer Vision for ADAS on Tensilica Vision P6

- Convolutional Neural Network (CNN) benchmark
  - Speedup of 64 through VP6 optimization
  - VP6: 330 CNNs / s @ 750 MHz, *4 W*
  - Core-i5: 83 CNNs/s @ 2,6 GHz, *45 W*
- 360° Surround View
  - VP6: 30 fps @ 750 MHz, *~4 W*
  - Intel GPP: 8 fps @ 4x3,4 GHz, *~65 W*



➢ Tensilica Vision P6 architecture (512b SIMD, 5-way VLIW, DMA, Gather-Scatter) fits to common Computer Vision use cases

Tensilica Vision P6 ideally suitable for a variety of Computer Vision algorithms in ADAS