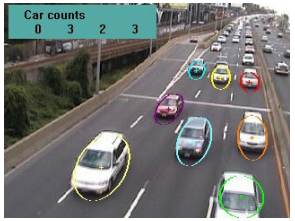
Energy-Efficient HOG-based Object Detection at 1080HD 60 fps with Multi-Scale Support

Amr Suleiman, Vivienne Sze

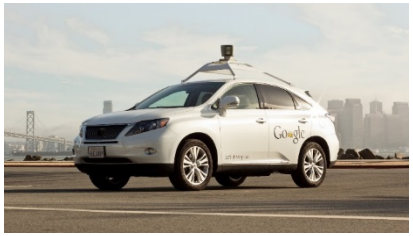
Massachusetts Institute of Technology

Motivation



Surveillance

Advanced Driver Assistance Systems (ADAS)



Self-driving cars

Unmanned Aerial Vehicles (UAV)

Mobile phones

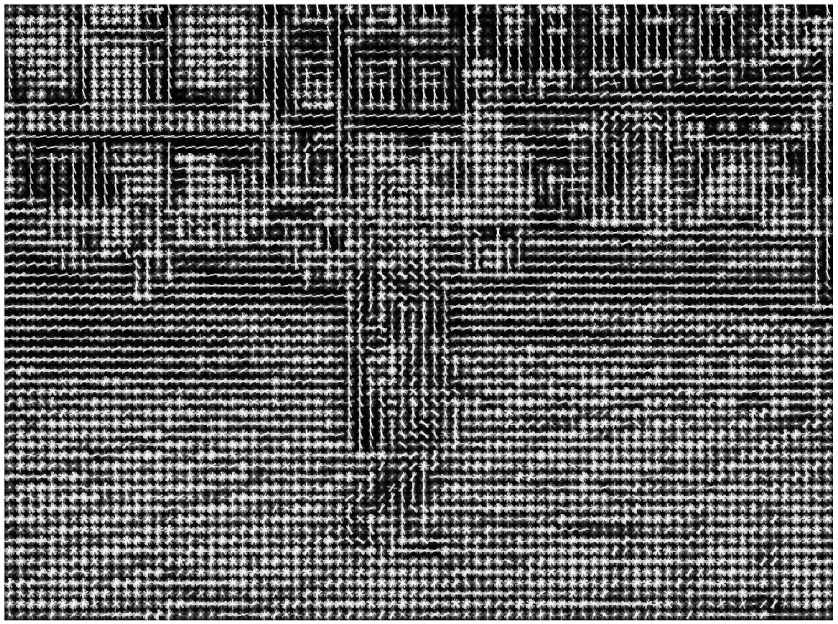
These applications require

- Real-time processing
- High frame rates
- High resolution images (1080HD)
- Low power

Dedicated ASIC design for object detectors

Overview of Object Detection Methodology (Sliding Window)

- Map the image into higher dimension feature space



Pixels

Features

Overview of Object Detection Methodology (Sliding Window)

- Map the image into higher dimension feature space.
- Use sliding window to sweep the image
  - Match with a pre-trained template

Pixels

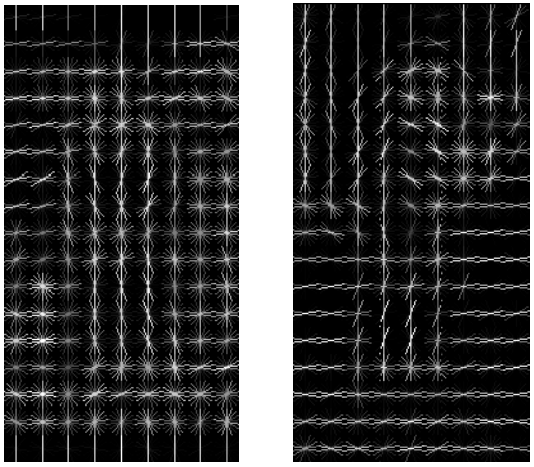
Features

Feature: HOG

- Histogram of Oriented Gradients (HOG) [Dalal & Triggs, *CVPR 2005*]

Pedestrian images

Corresponding HOG features

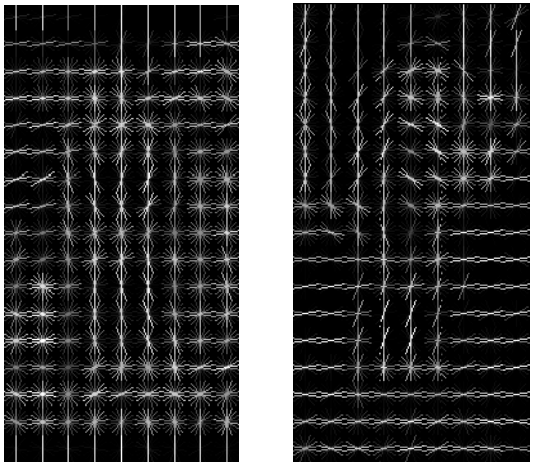


Feature: HOG

- Histogram of Oriented Gradients (HOG) [Dalal & Triggs, *CVPR 2005*]

Pedestrian images

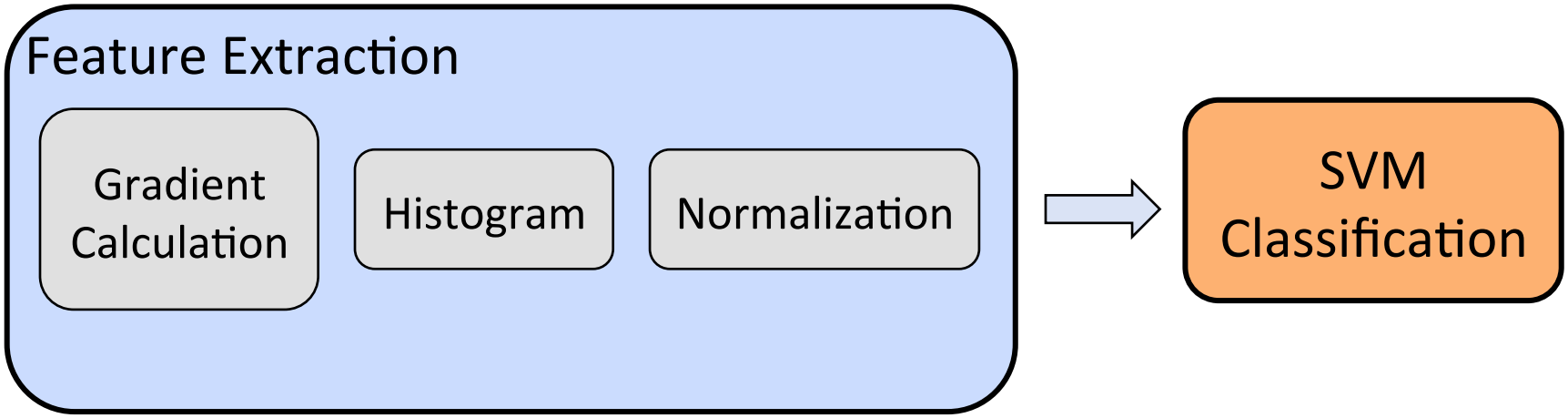
Corresponding HOG features



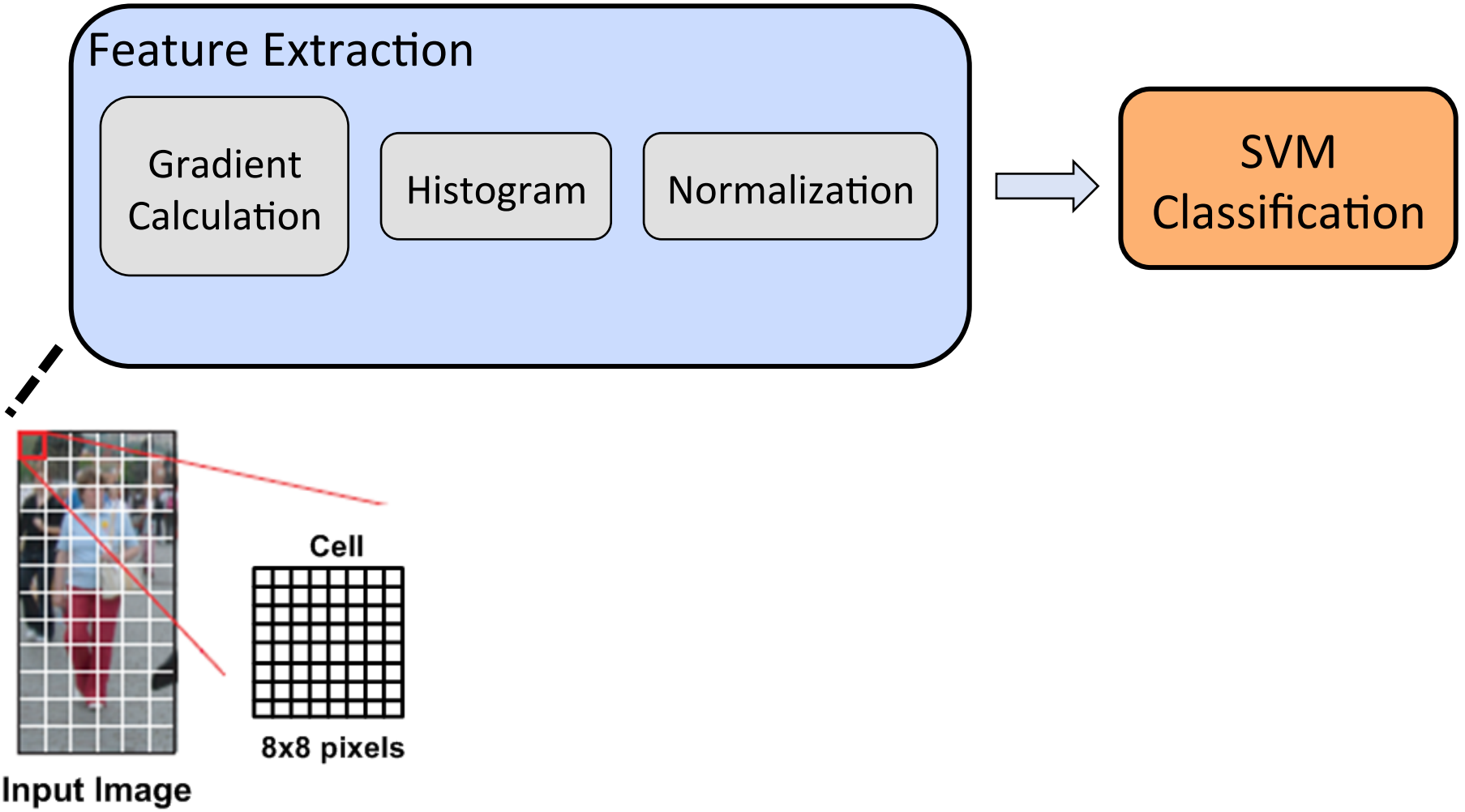
Training

SVM Template

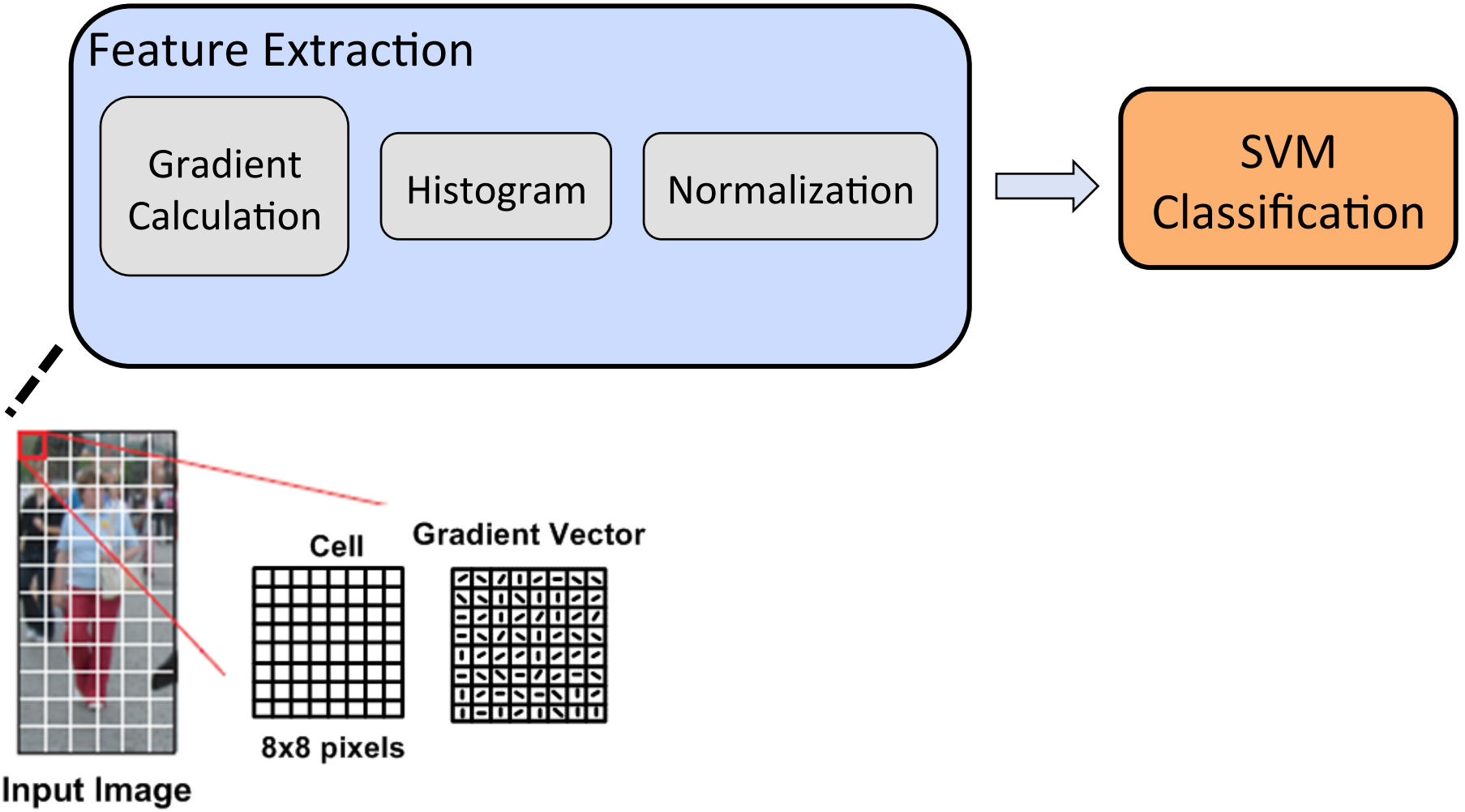
HOG-based Detection Pipeline



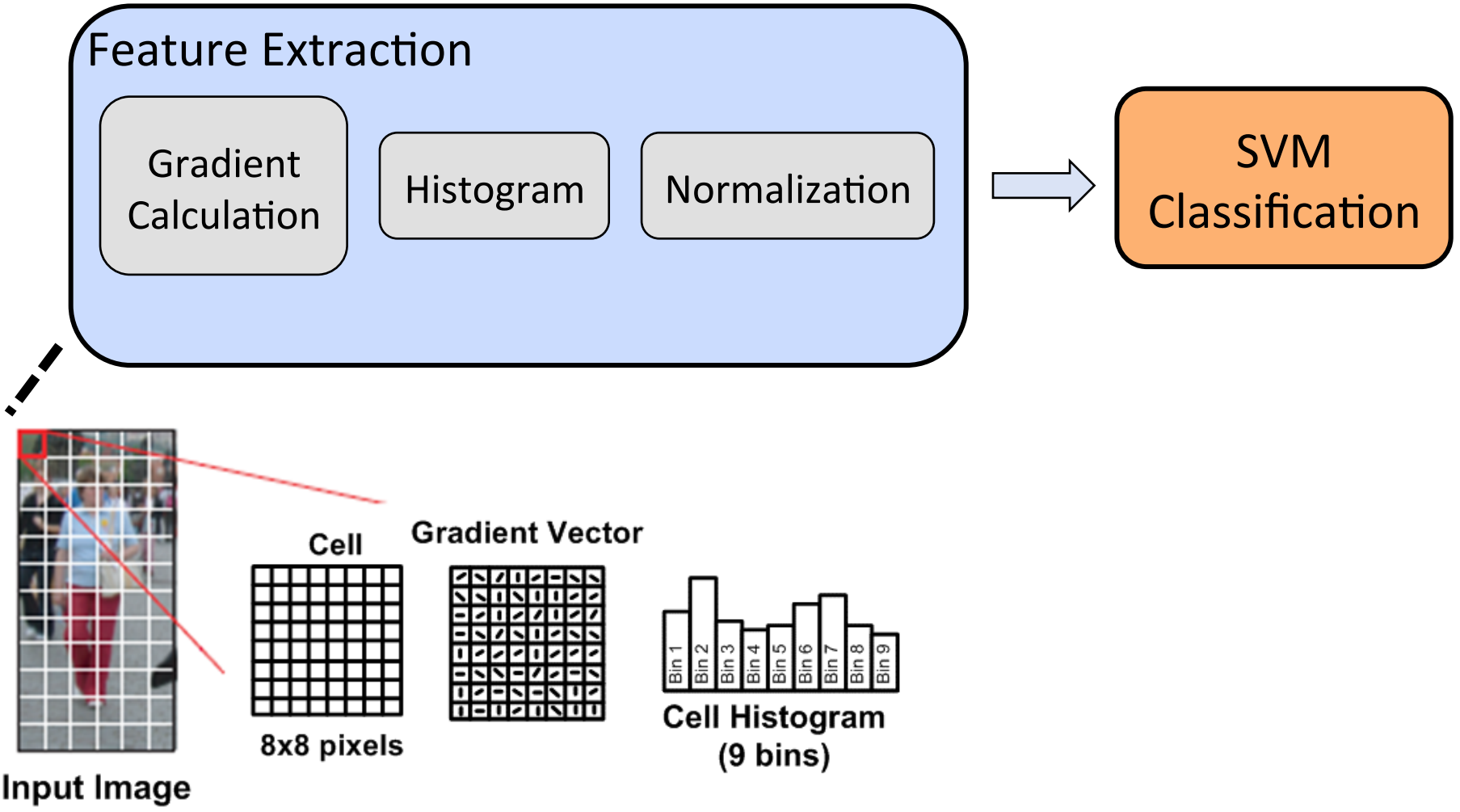
HOG-based Detection Pipeline



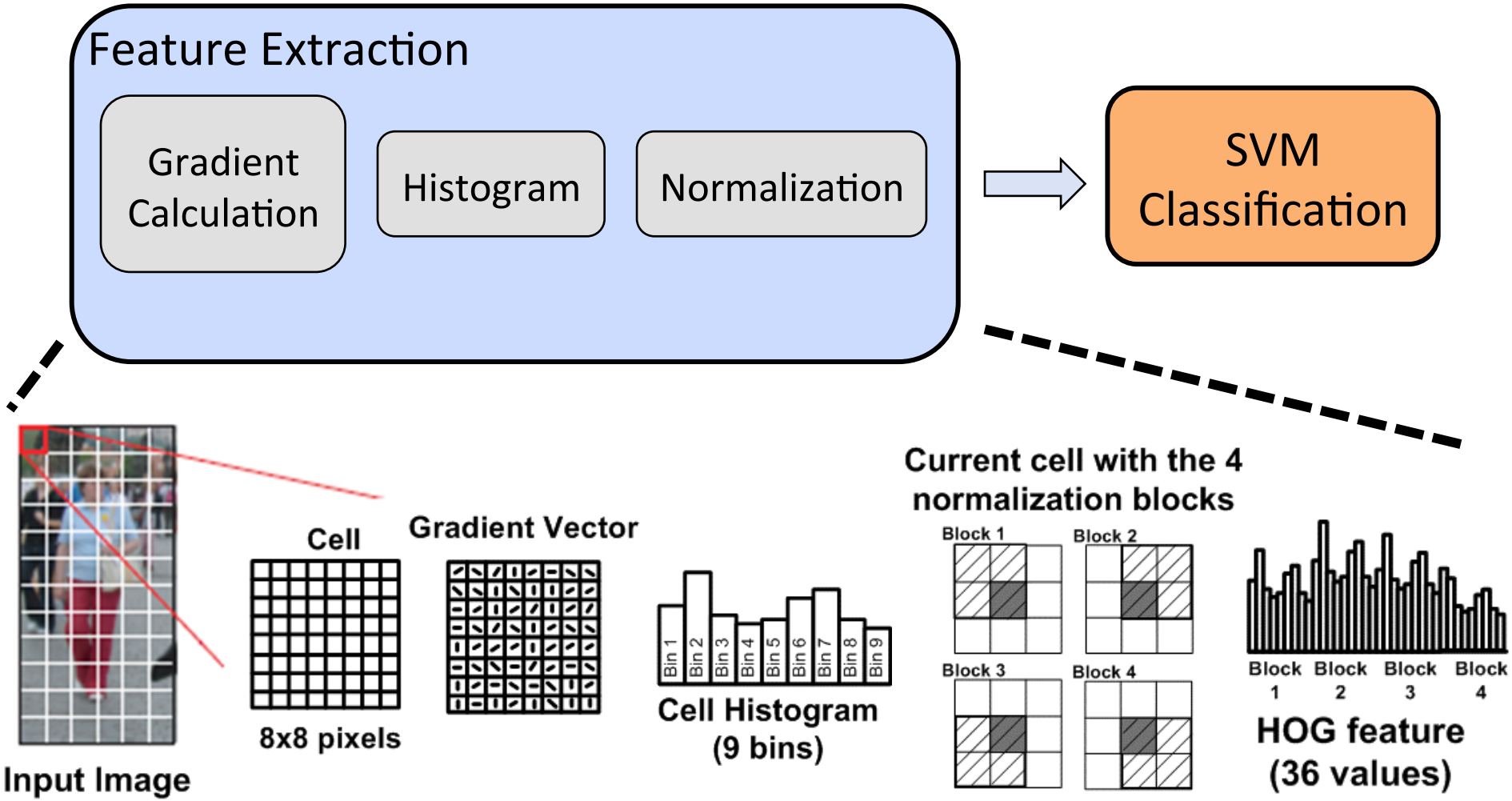
HOG-based Detection Pipeline



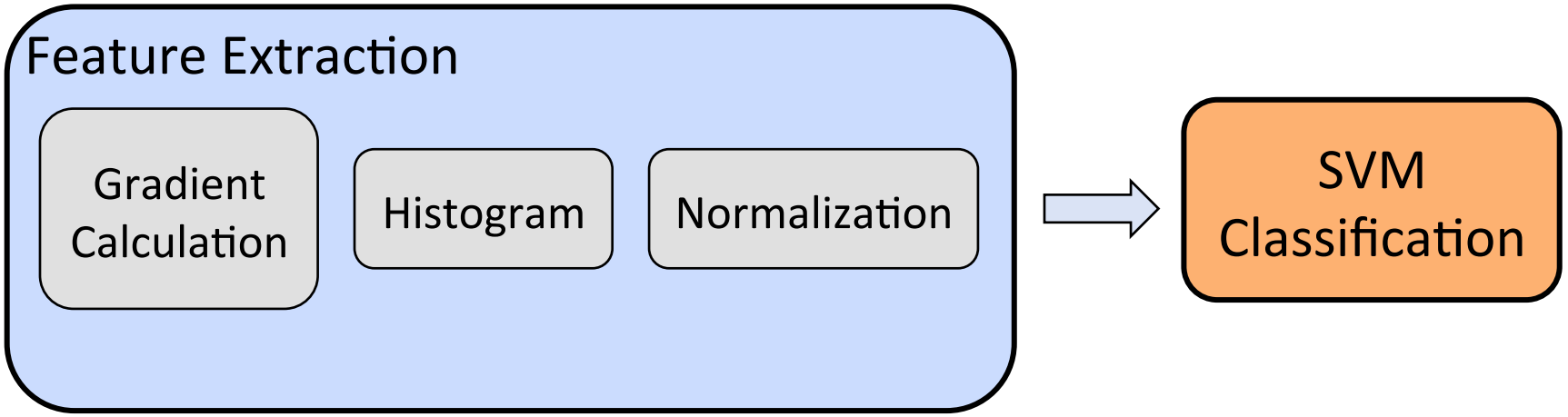
HOG-based Detection Pipeline

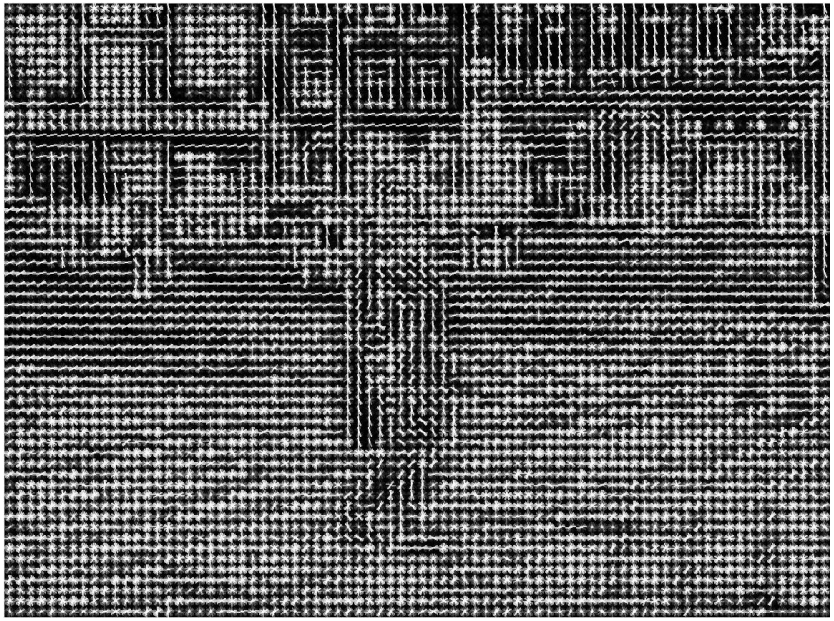


HOG-based Detection Pipeline



HOG-based Detection Pipeline





HOG-based Detection Pipeline

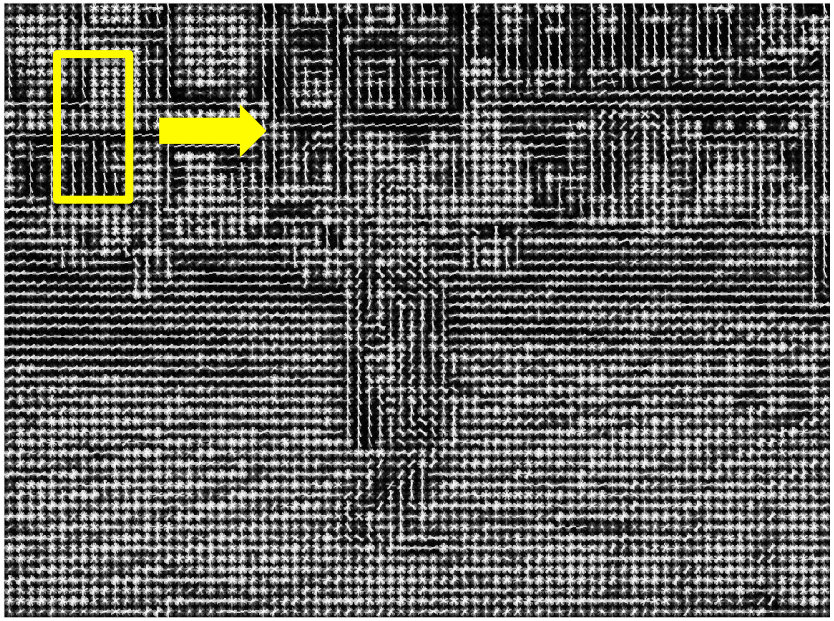
Feature Extraction

Gradient Calculation

Histogram

Normalization

SVM Classification



*

=> score { True, False }

Detecting Objects with Different Sizes

- One SVM template detects objects with one specific size

Single scale detection is not practical

Detecting Objects with Different Sizes

- Process different resolution of the same frame

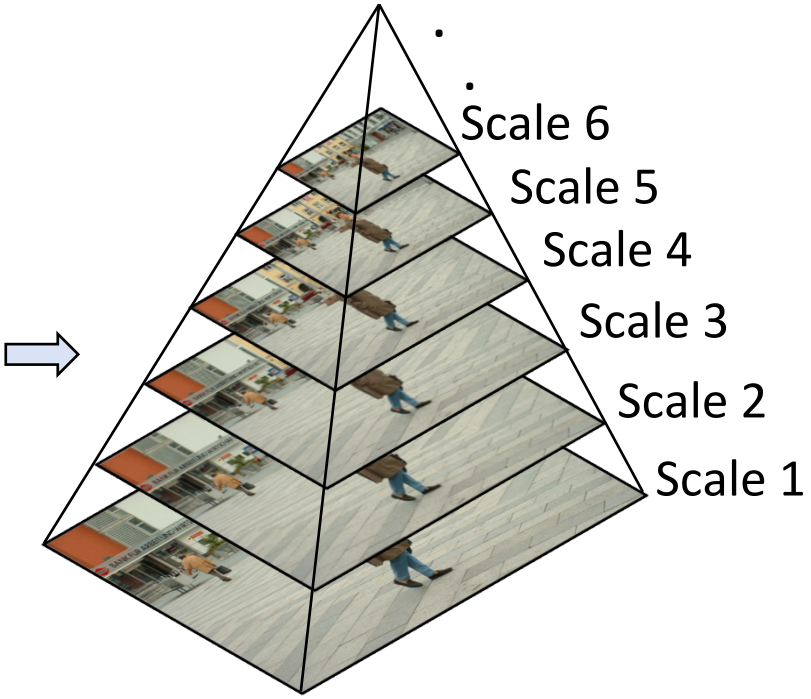


Image Pyramid

Detecting Objects with Different Sizes

- Process different resolution of the same frame.

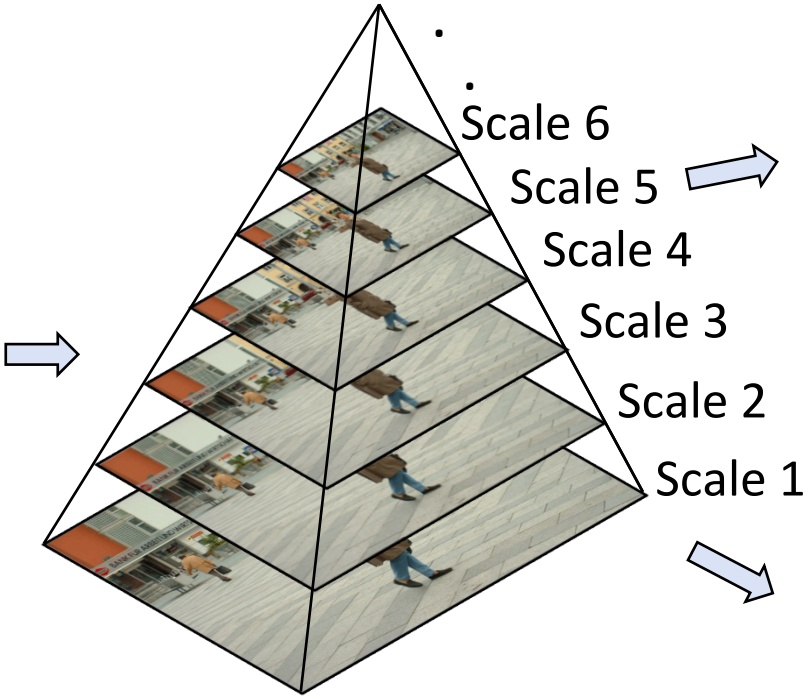


Image Pyramid

Main Contributions

- Multi-scale processing enables robust detection
  - However increases energy and memory requirements

Main Contributions

- Multi-scale processing enables robust detection
  - However increases energy and memory requirements

Approach

- Efficient scale selection and implementation of scales generator
- Parallel detectors with voltage scaling
- Image pre-processing to reduce multi-scale memory overhead

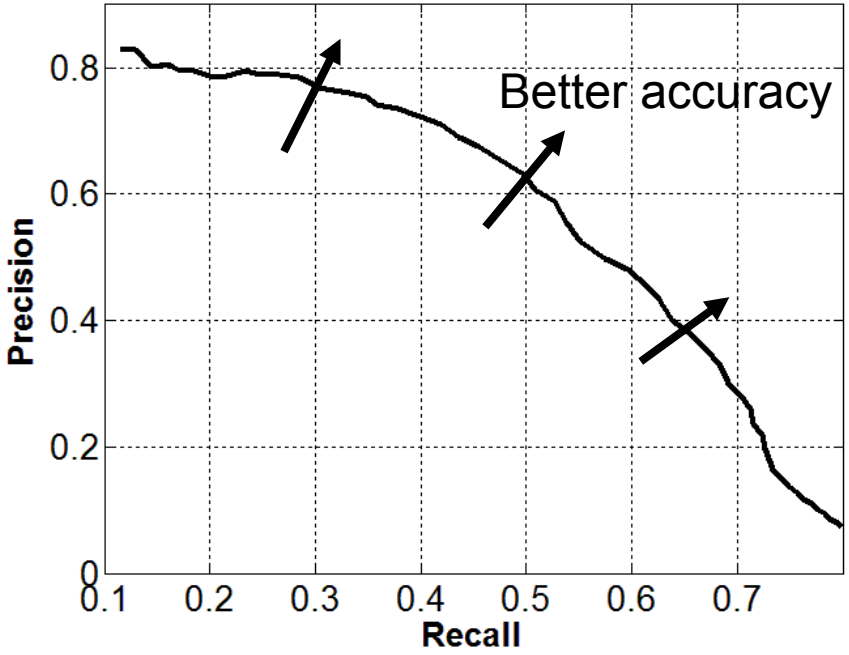
Detection Accuracy Measurement

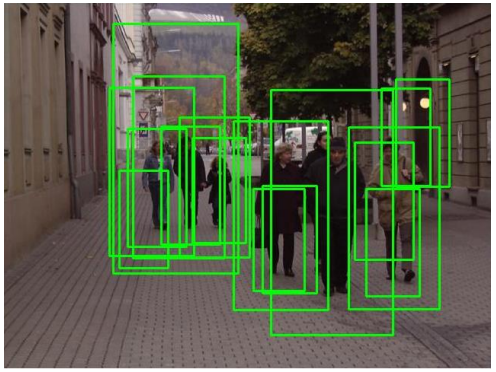
- Measuring accuracy by Precision-Recall curves

  Precision: Fraction of detected windows that are true objects

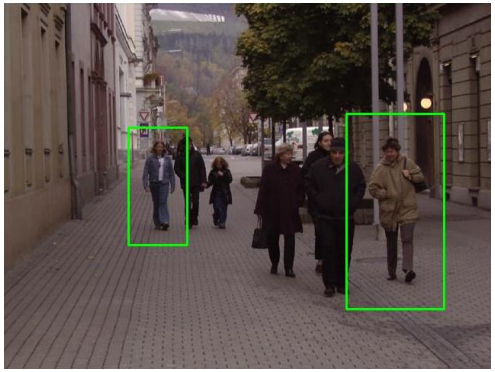
  Recall: Fraction of true objects that are detected

- Average precision (**AP**) = area under curve





Low Precision
High Recall

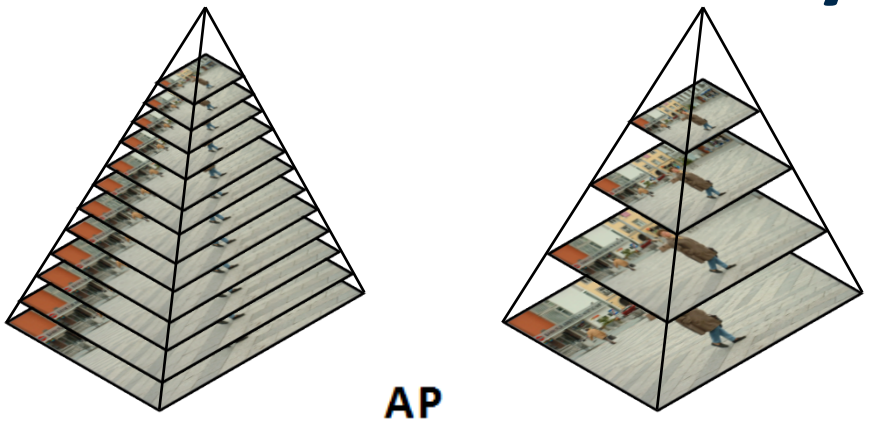


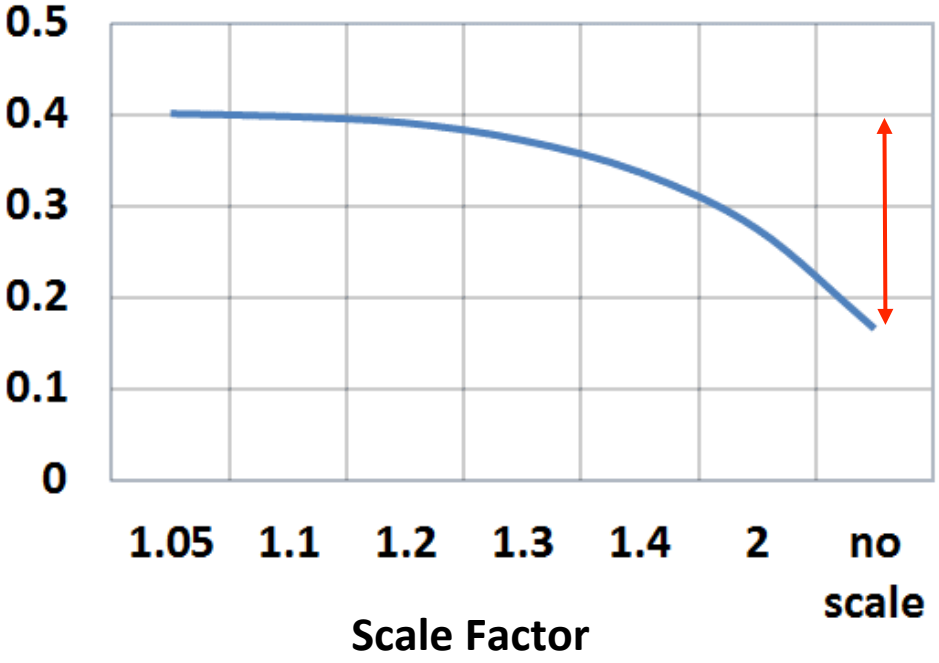
High Precision
Low Recall

Main Contributions

- Efficient scale selection and implementation of scales generator
- Parallel detectors with voltage scaling
- Image pre-processing to reduce multi-scale memory overhead

Detection Accuracy vs. Complexity



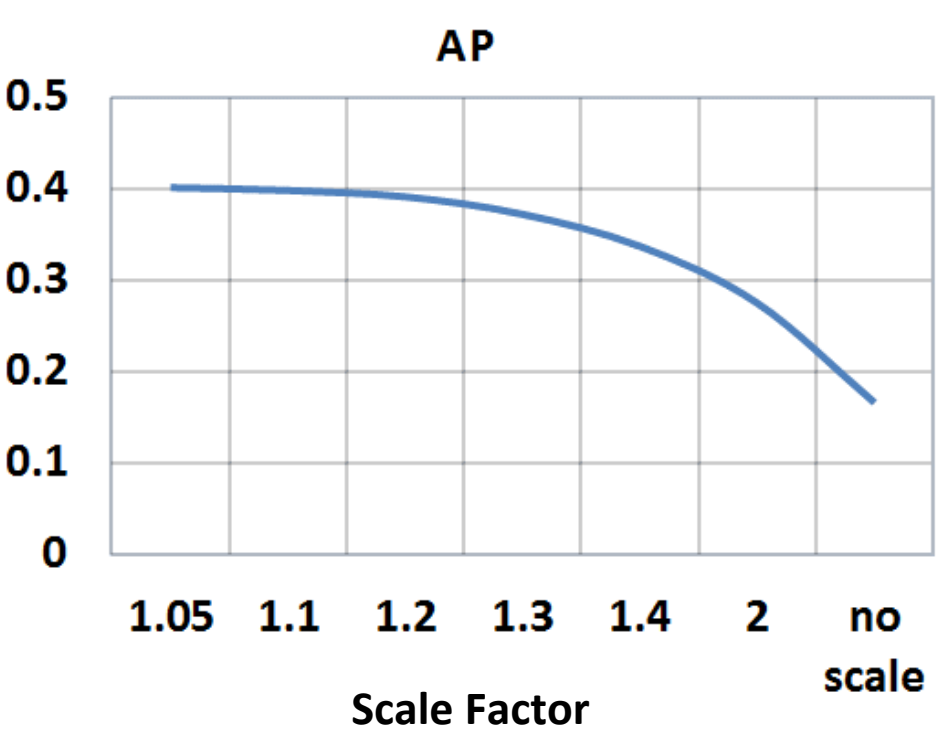


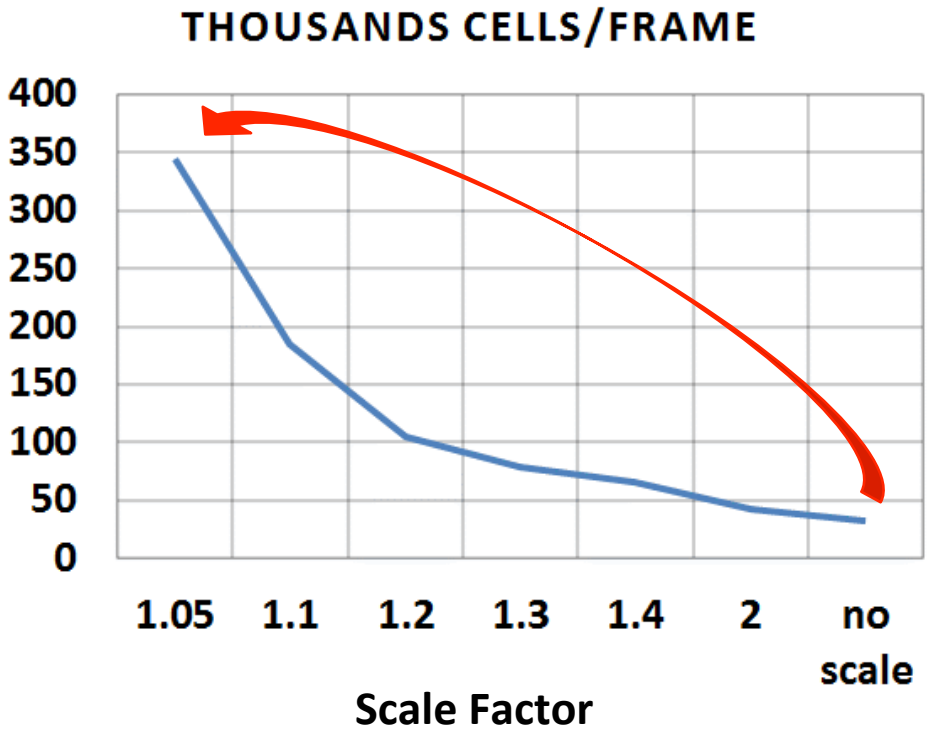
Multi-scale increases AP by **2.4x**

Using INRIA person dataset

Detection Accuracy vs. Complexity

Multi-scale results in **10x** more cells to process



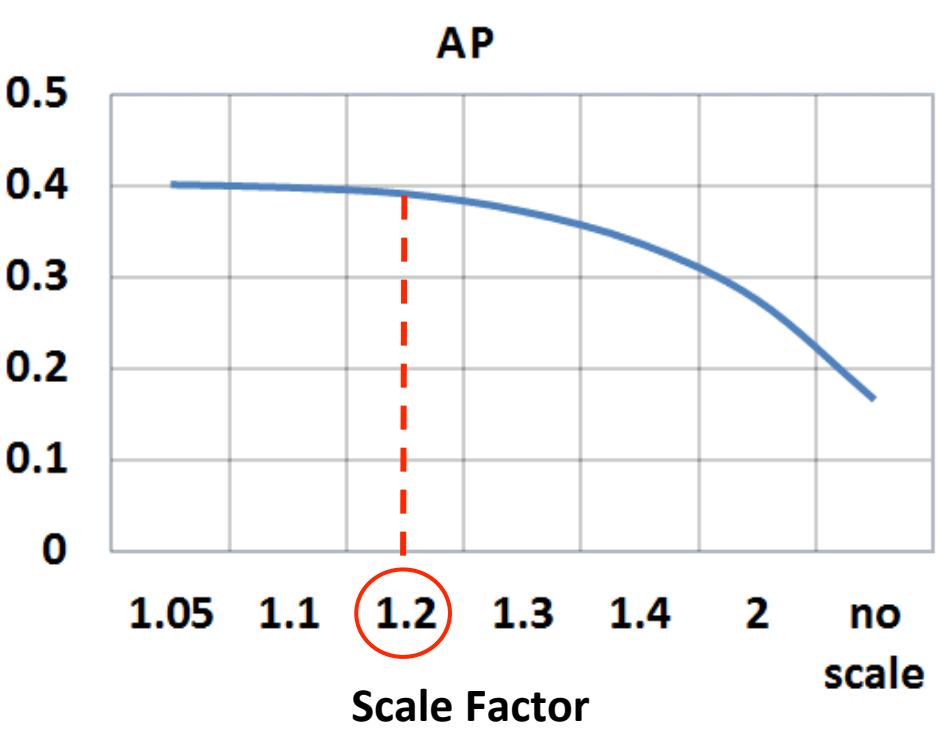


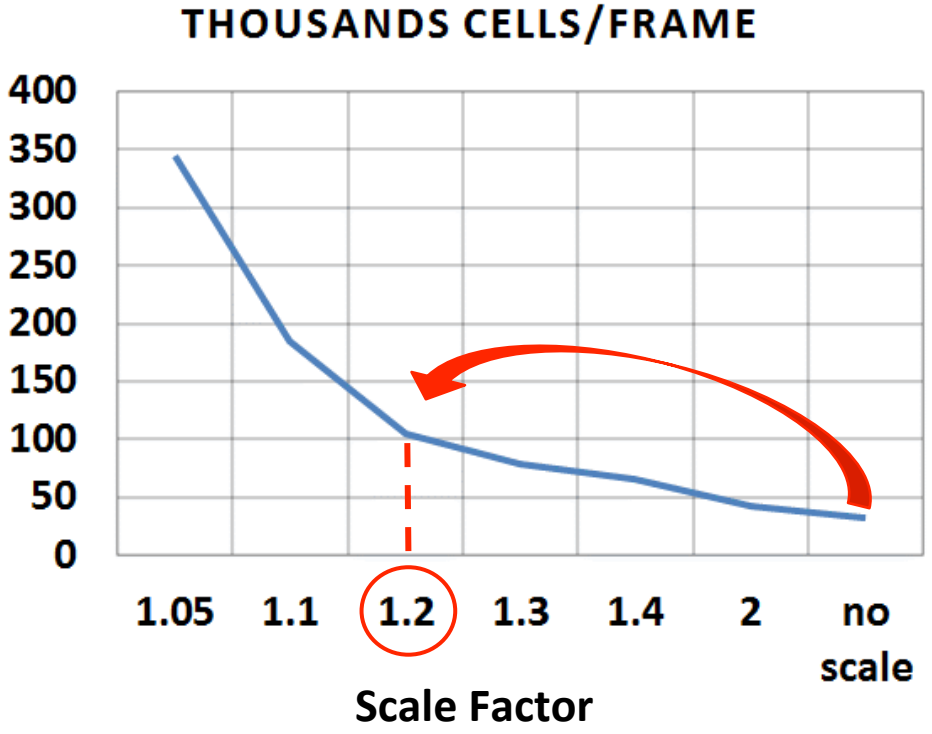
Using INRIA person dataset

Detection Accuracy vs. Complexity

A scale factor of 1.2 gives:

- **3.3x** increase in number of cells (rather than 10x)
- Only **0.01 AP** loss

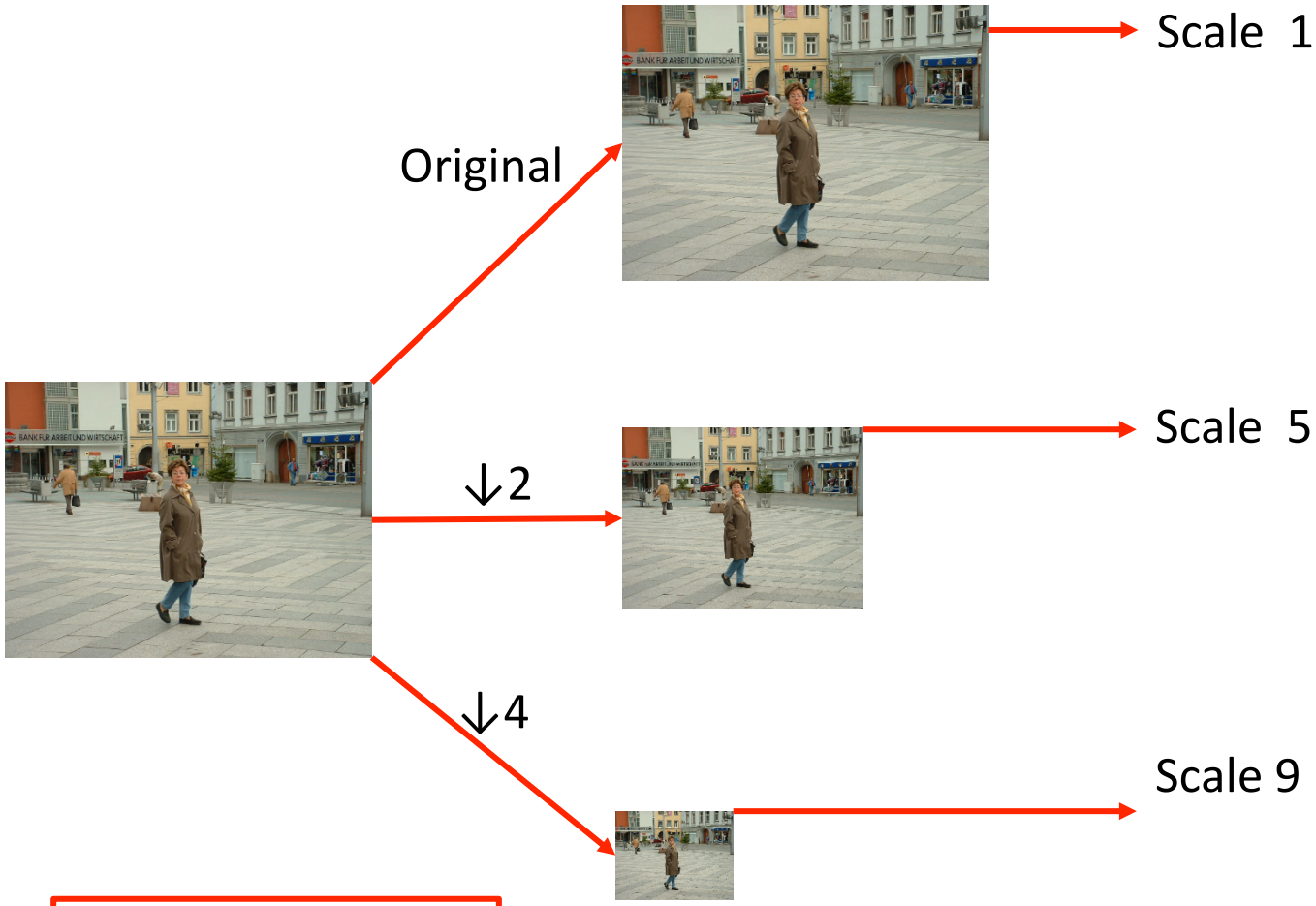




Using INRIA person dataset

Scales Generation

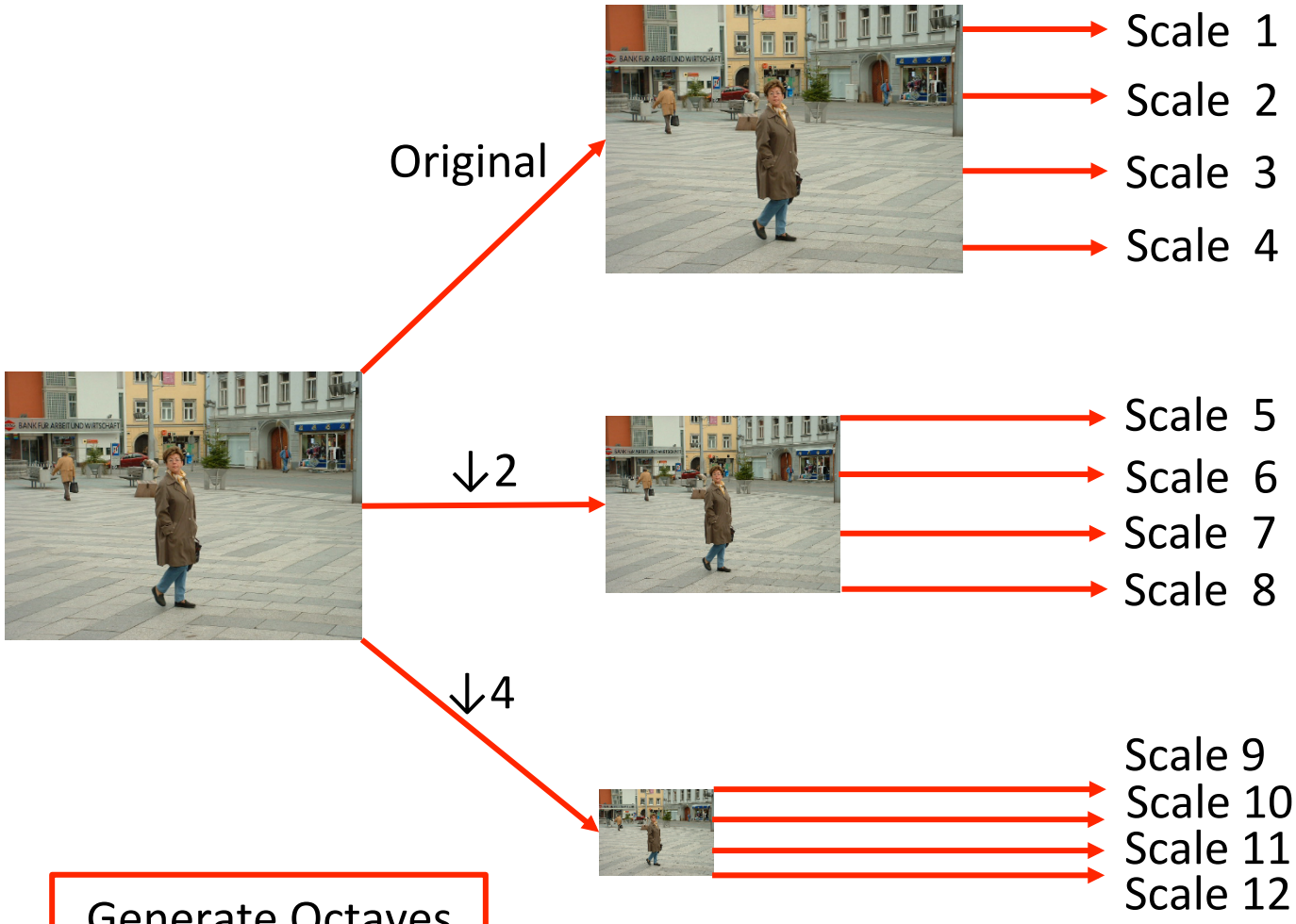
Image pyramid with 12 scales generated per frame



Generate Octaves

Scales Generation

Image pyramid with 12 scales generated per frame

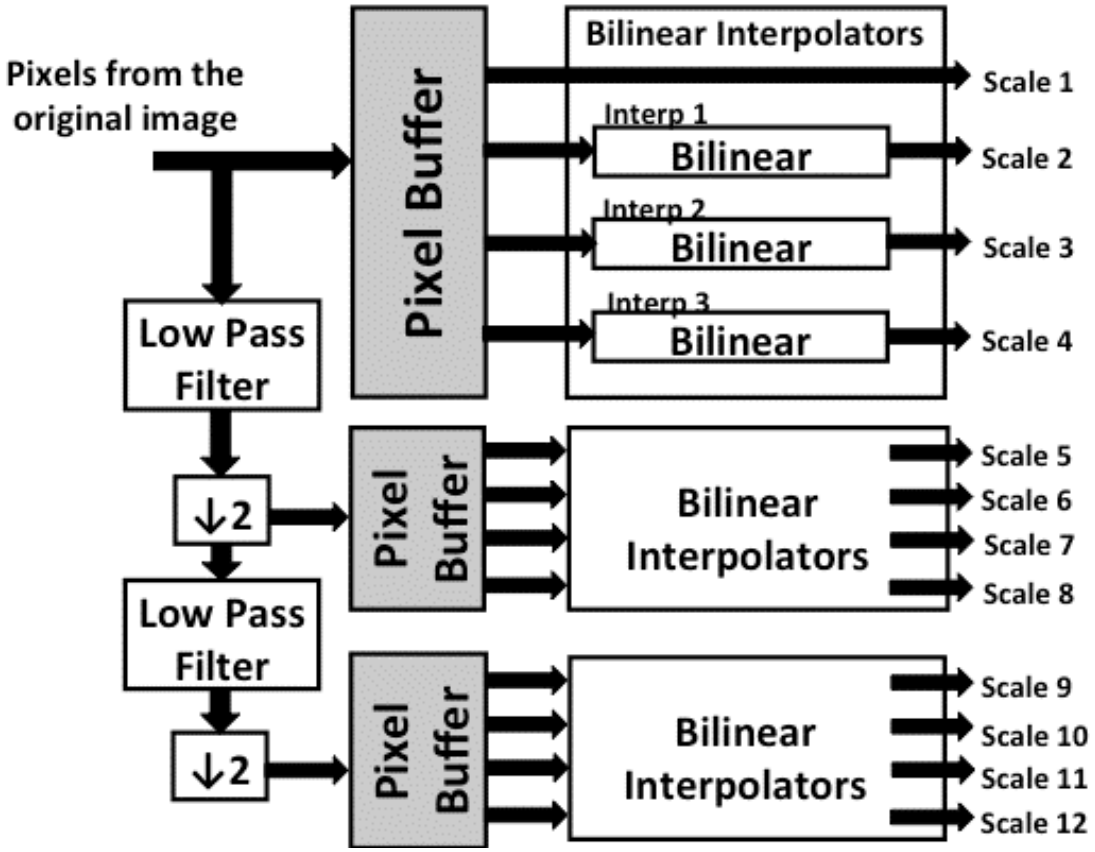


Generate Octaves

Bilinear Interpolation

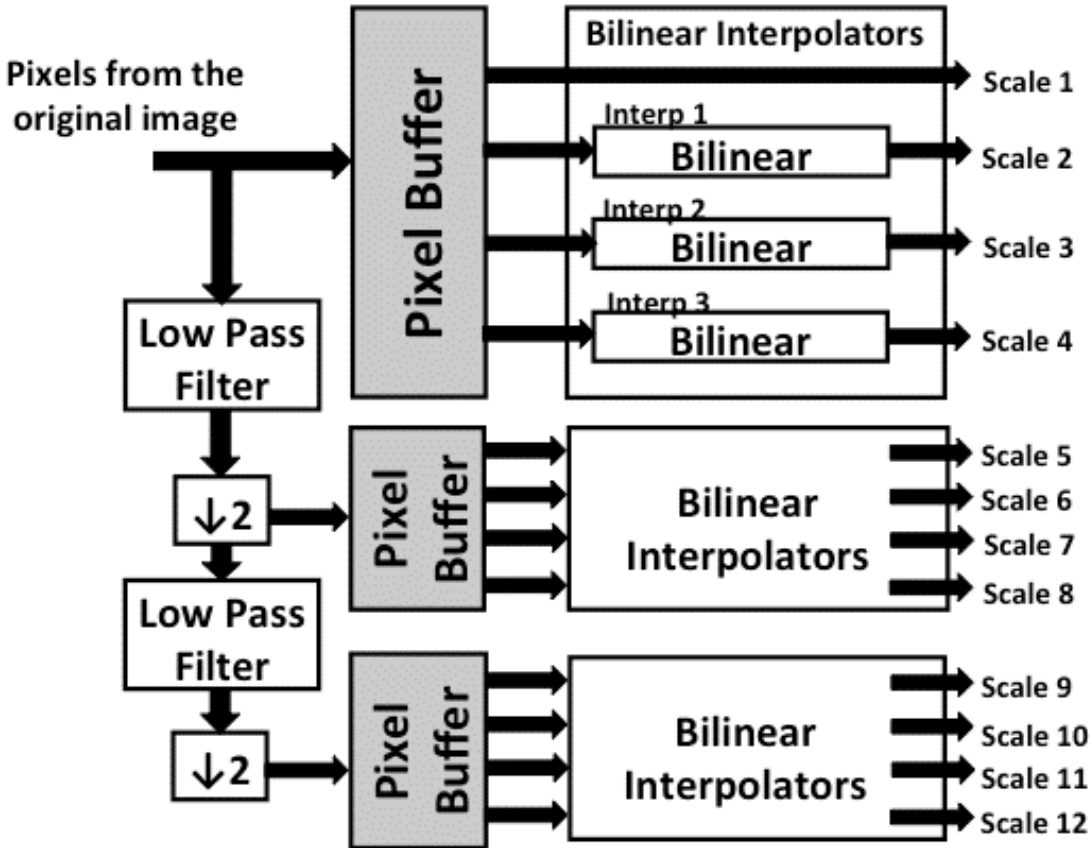
Scale Generator Architecture

- Three pixel buffers to support interpolation
- On-the-fly processing to reduce memory requirements



Scale Generator Architecture

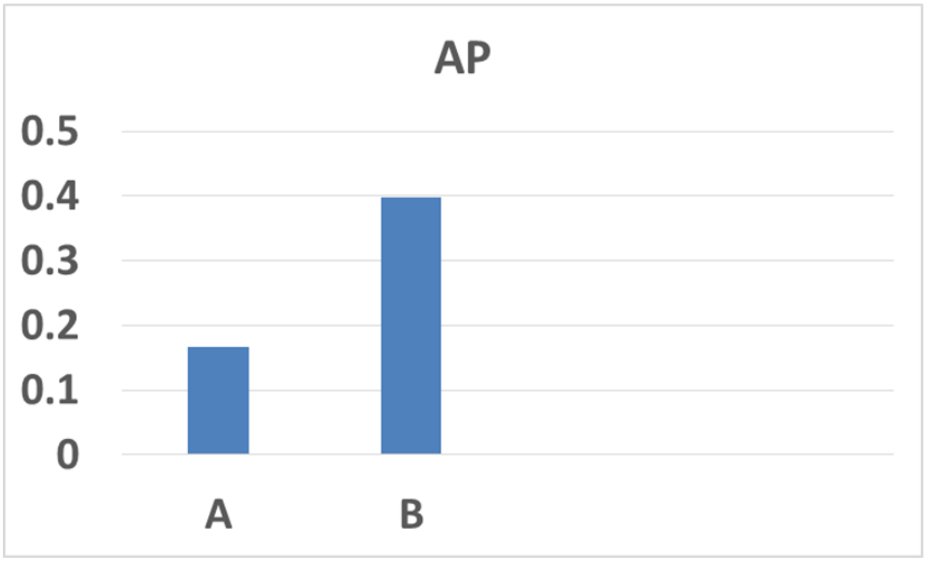
- Three pixel buffers to support interpolation.
- On-the-fly processing to reduce memory requirements.

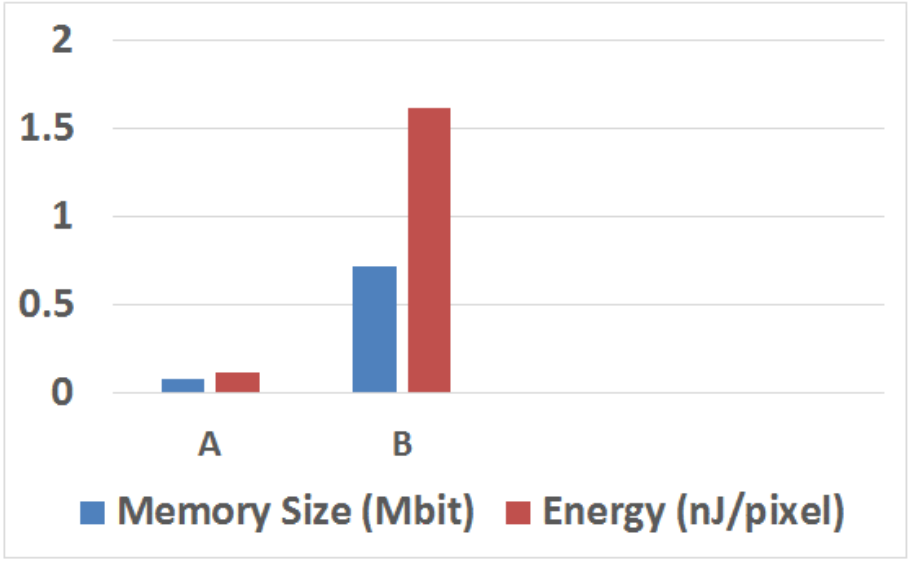


No scales are stored on-chip

Multi-Scale Detection Summary

- **Detection:** Improved by **2.4x**
- **Energy:** Increased by **14x**
- **Memory:** Increased by **8.8x**





A: Single scale, single detector
B: Multi-scale, single detector

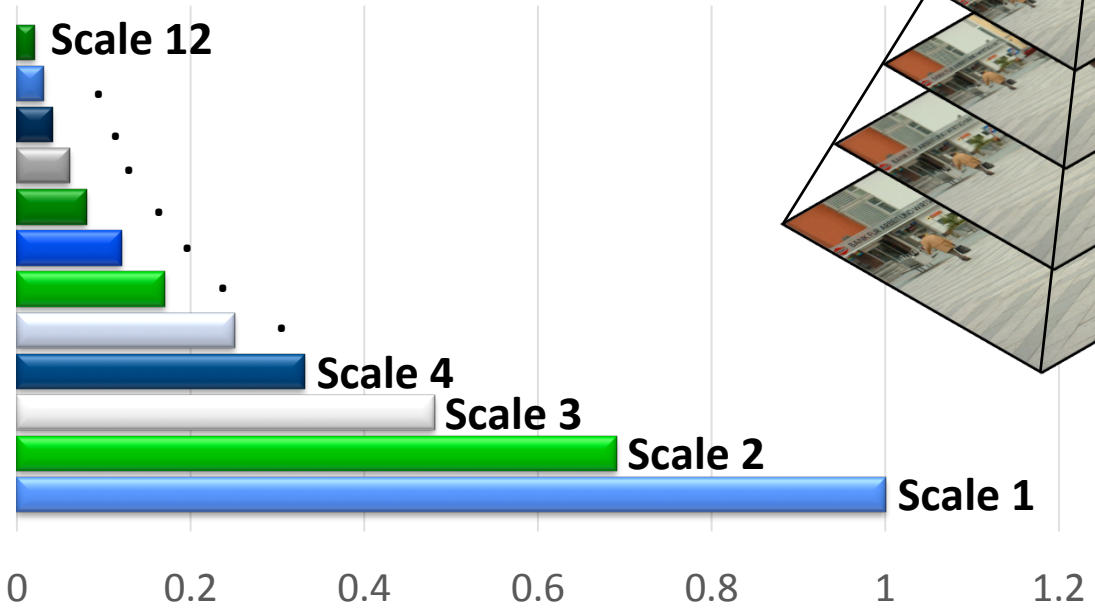
Constant throughput of 60 fps

Main Contributions

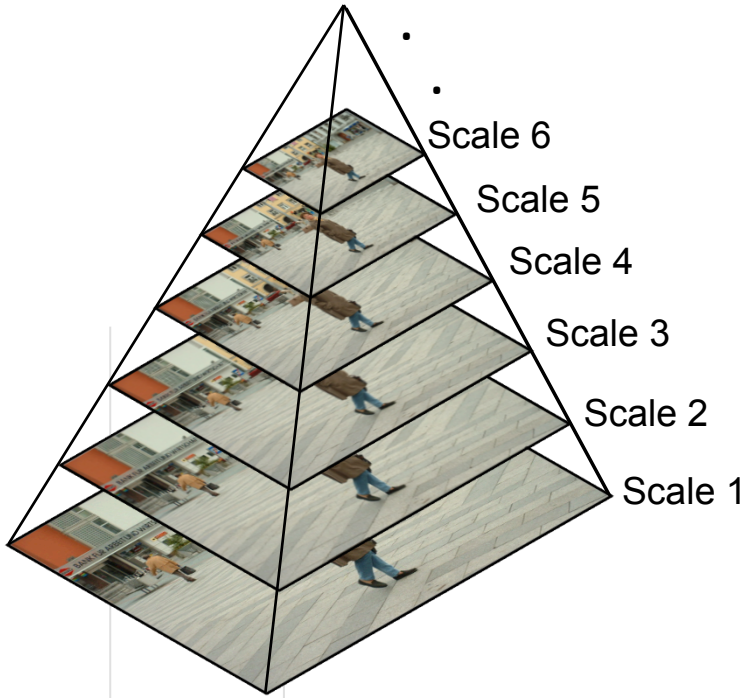
- Efficient scale selection and implementation of scales generator
- Parallel detectors with voltage scaling
- Image pre-processing to reduce multi-scale memory overhead

Scales Workload

Scale Generator

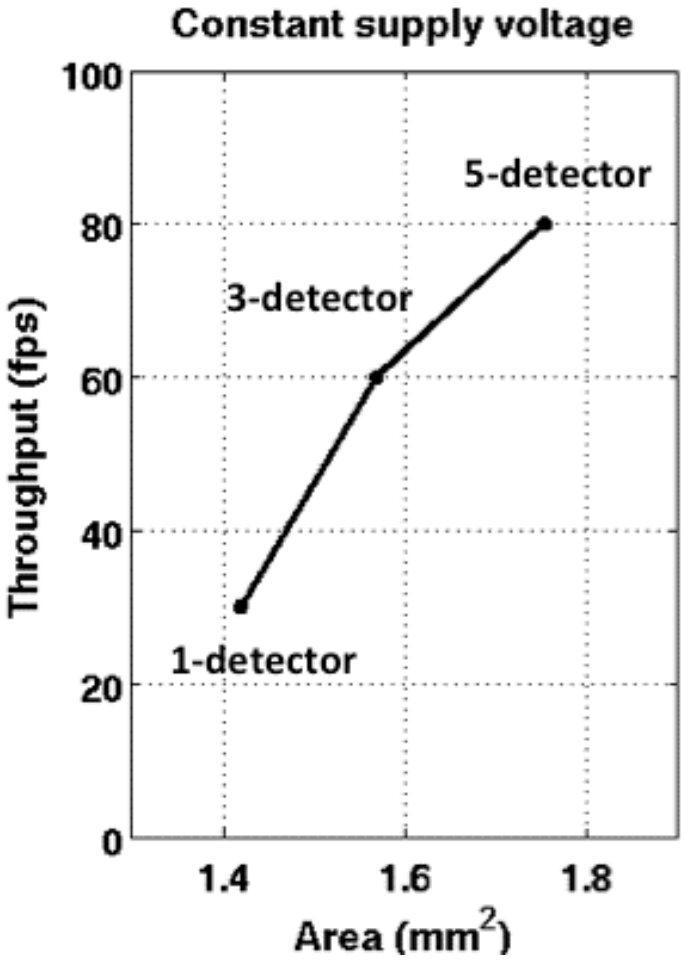


Normalized number of pixels per scale

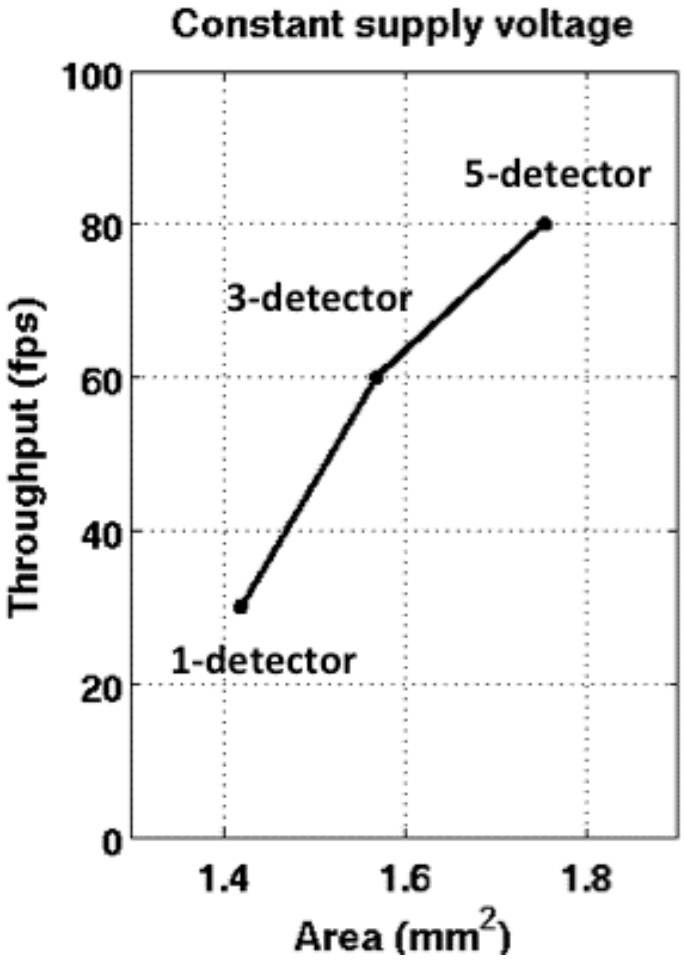


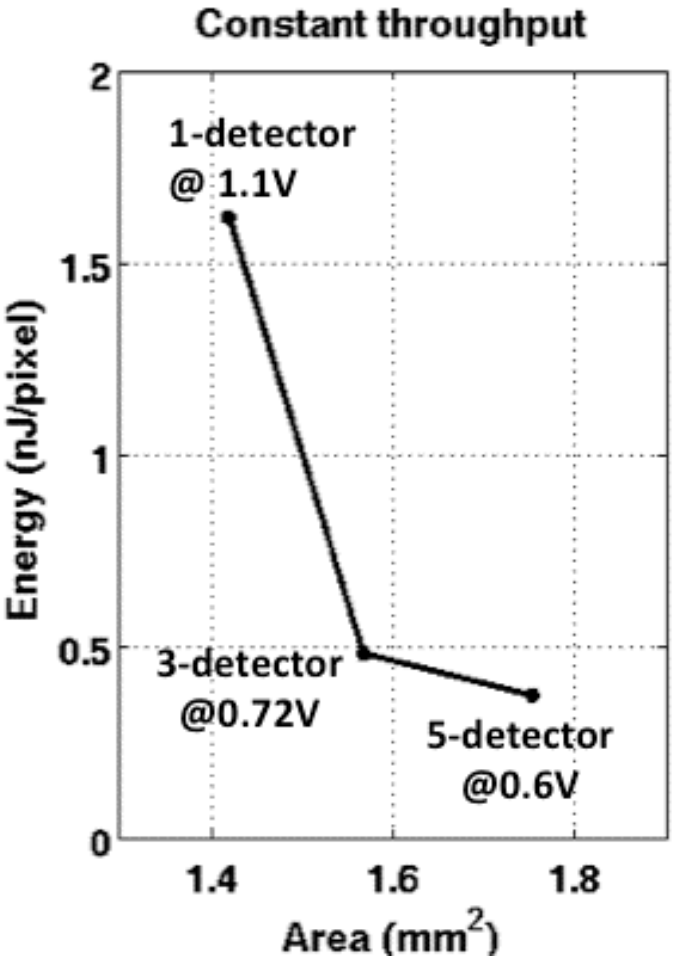
Distribute workload equally across parallel detectors

Parallel Detectors

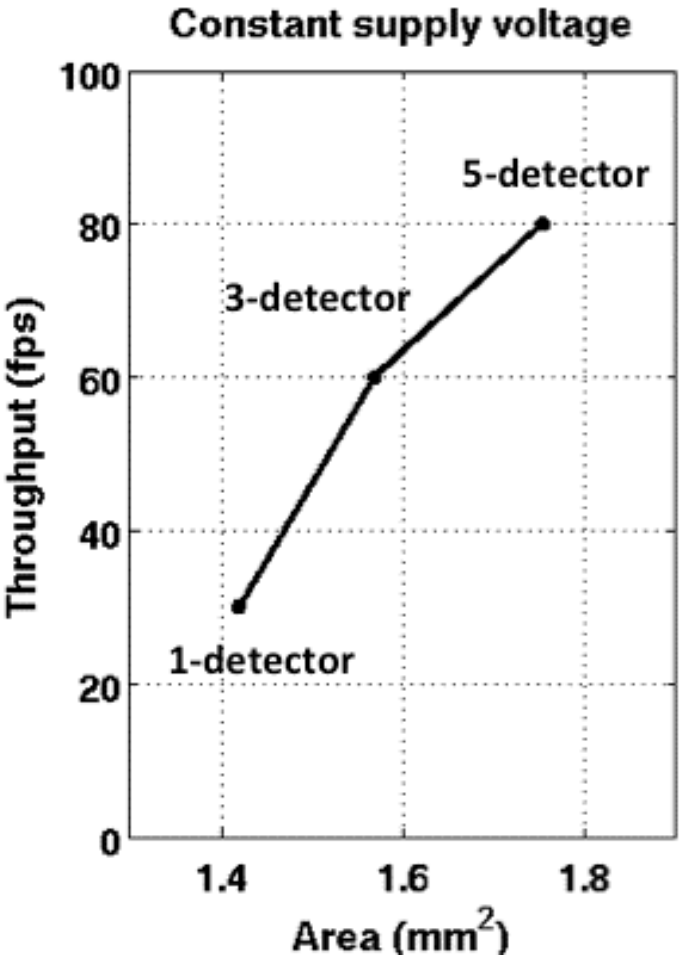


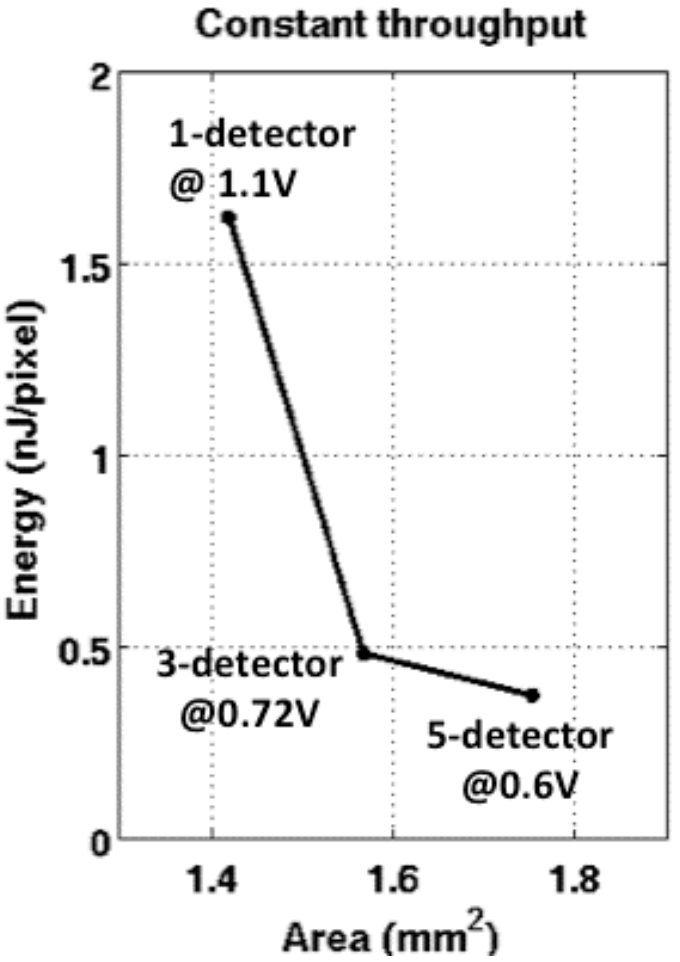
Parallel Detectors





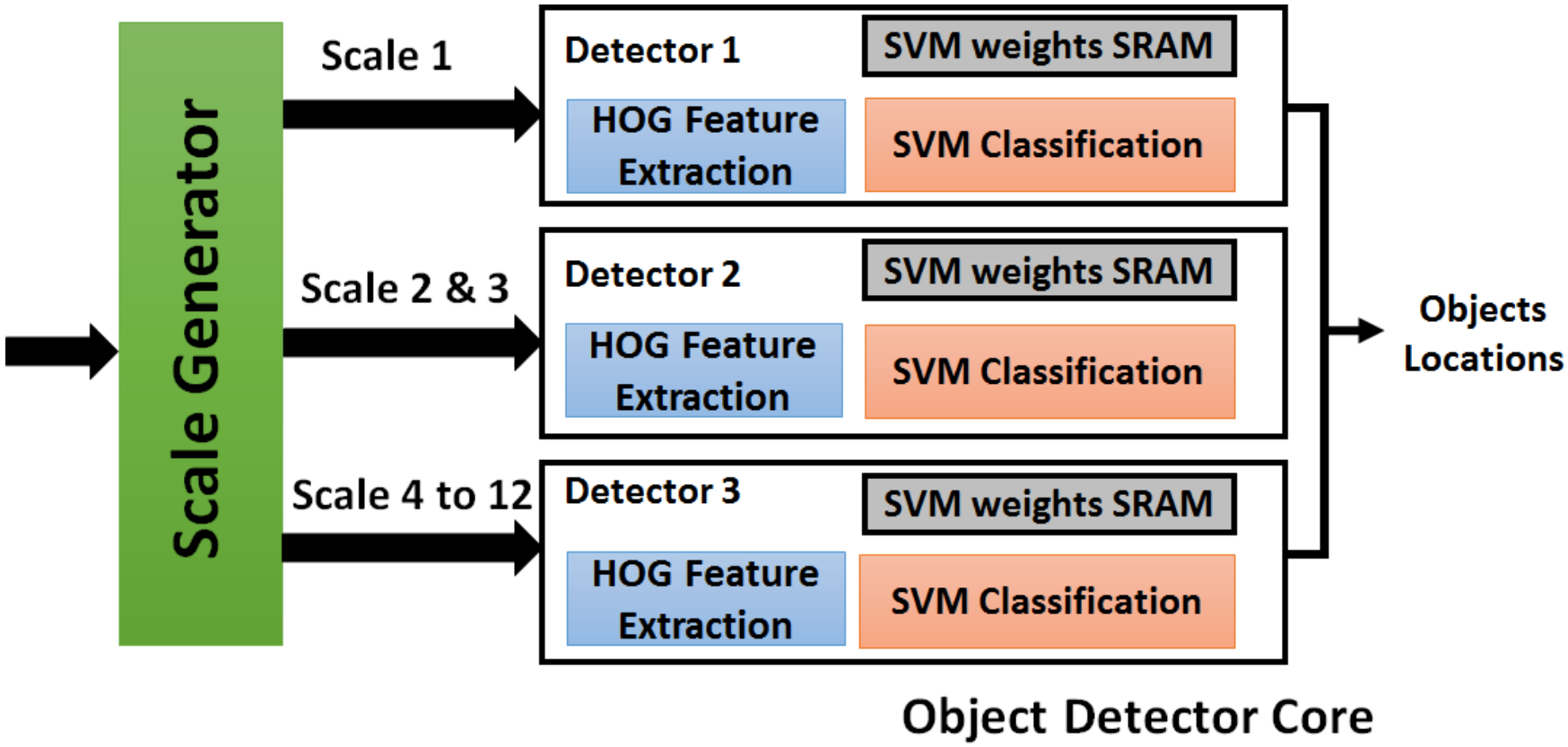
Parallel Detectors



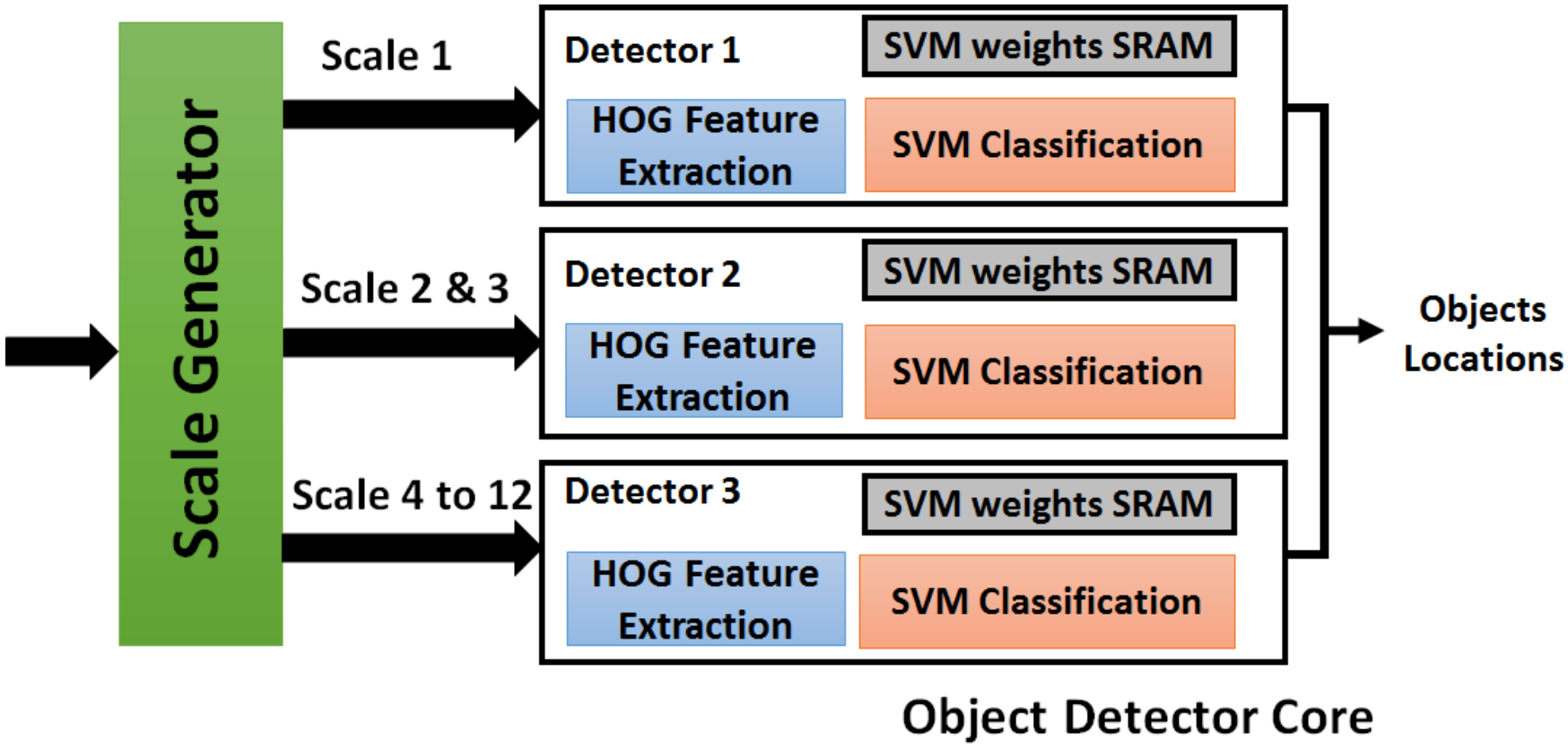


3-detector configuration is chosen

HOG-based Object Detection System

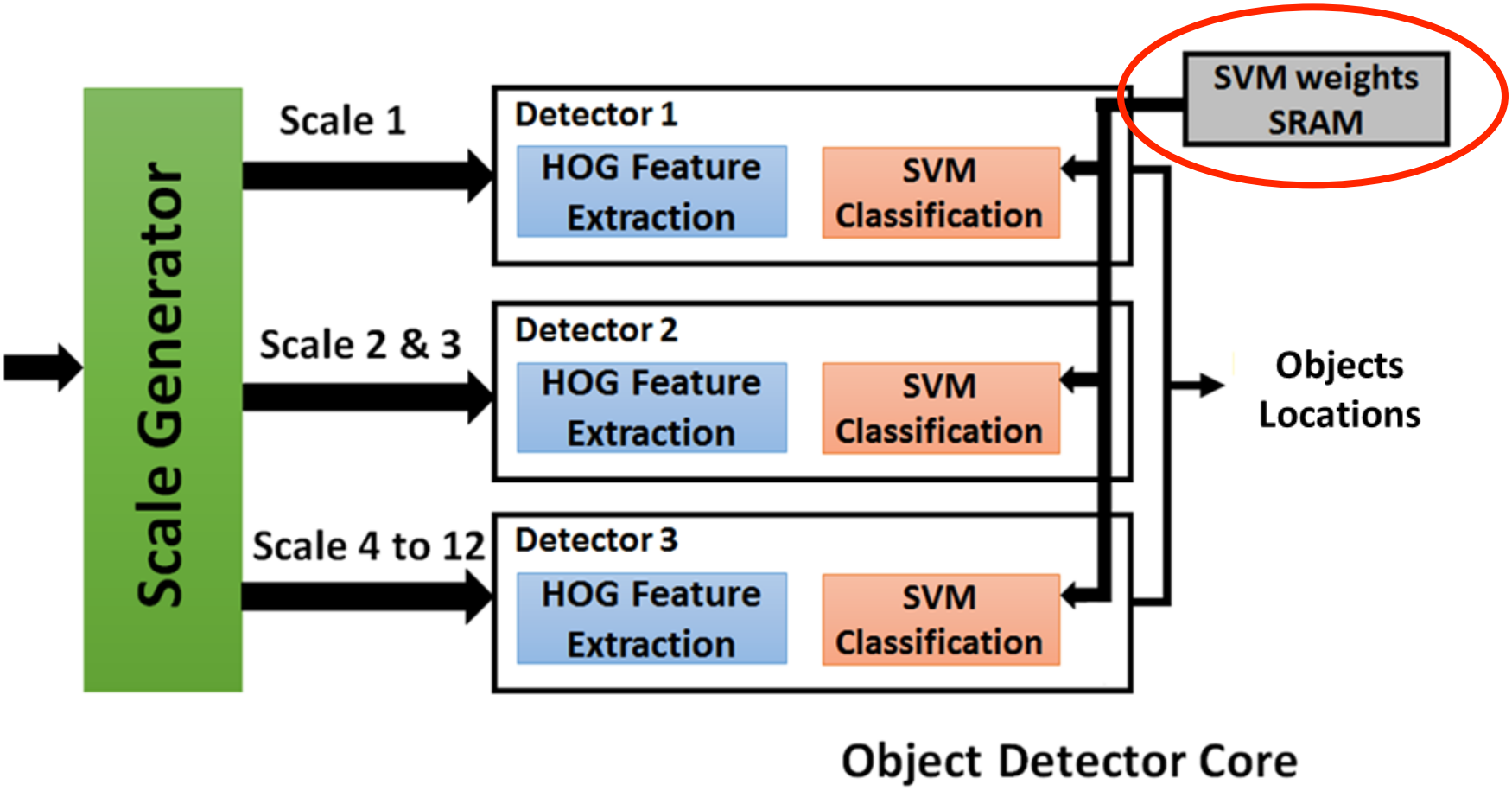


HOG-based Object Detection System



SVM weights SRAM consumes **37%** of a single detector power

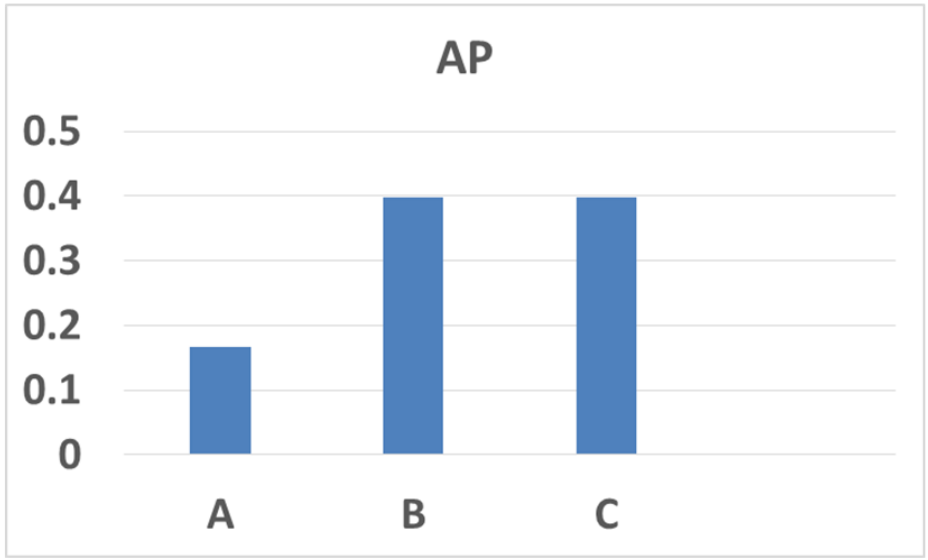
HOG-based Object Detection System

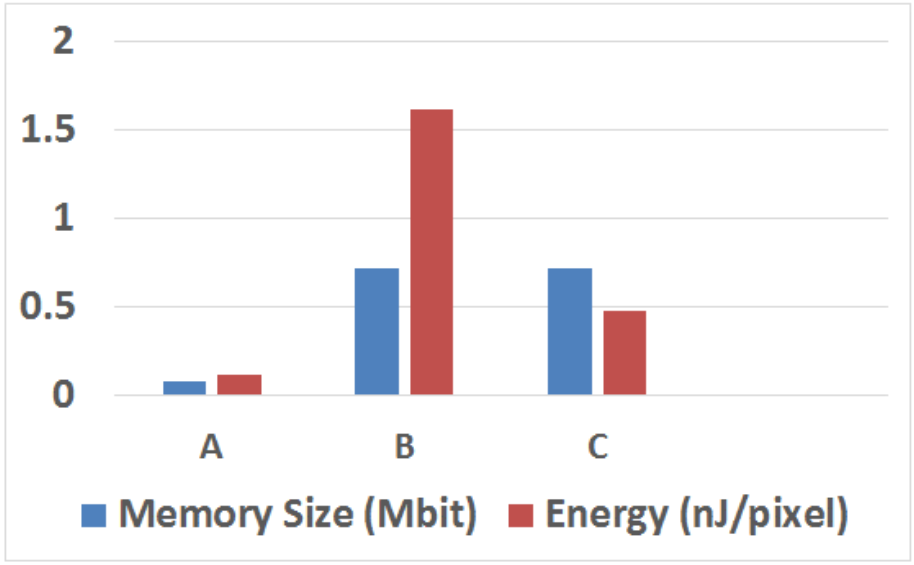


Synchronize detectors to share SVM weight memory
(**20%** reduction in power)

Parallelism and Voltage Scaling Summary

- **Detection:** Not affected
- **Energy:** Reduced by **3.4x**
- **Memory:** Not affected





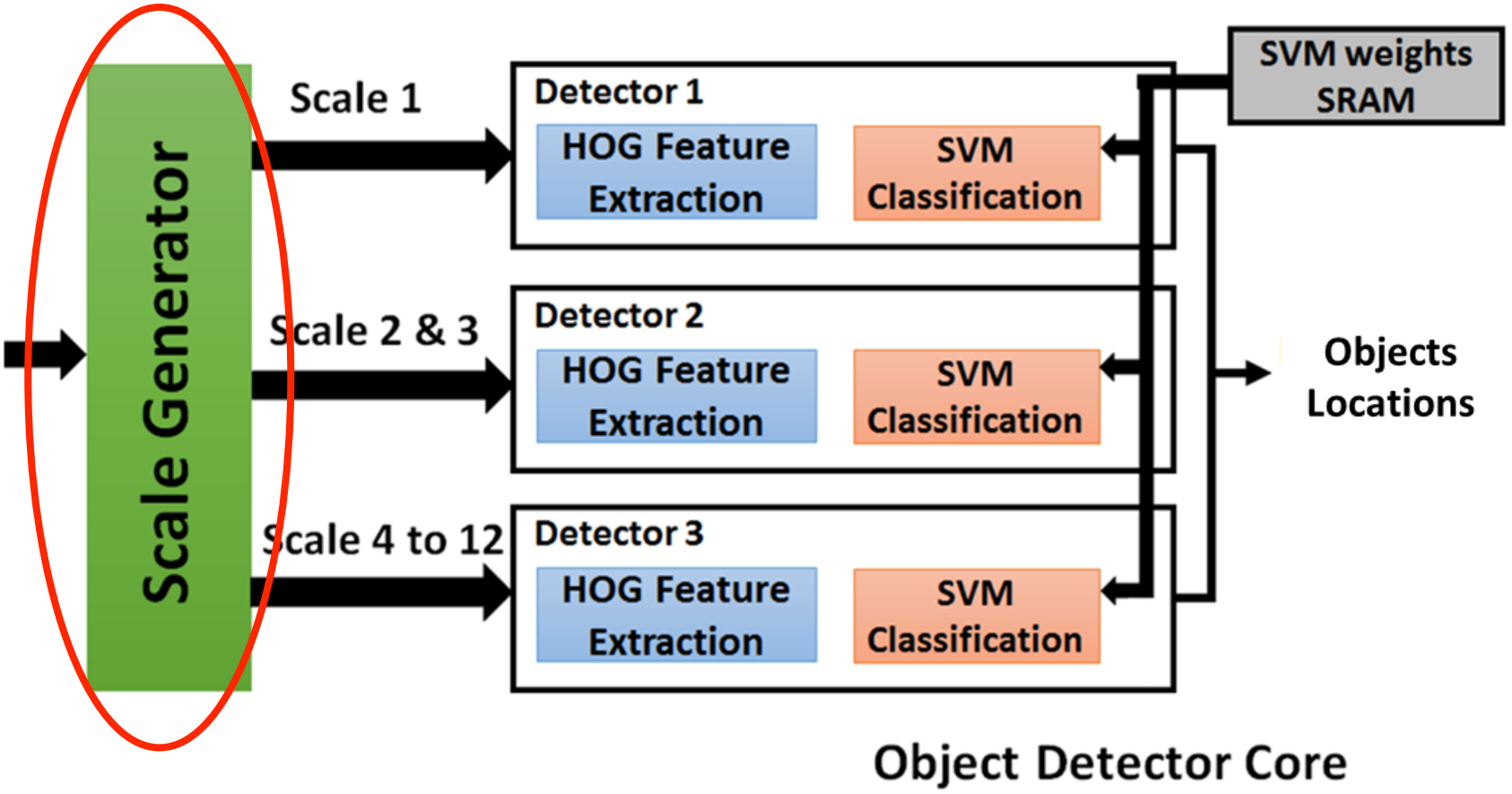
A: Single scale, single detector
B: Multi-scale, single detector
C: Multi-scale, 3-detectors

Constant throughput of 60 fps

Main Contributions

- Efficient scale selection and implementation of scales generator
- Parallel detectors with voltage scaling
- Image pre-processing to reduce multi-scale memory overhead

HOG-based Object Detection System



Scale generator has half of the system memory

Image Pre-Processing

- Pre-process to reduce scales generation overhead.

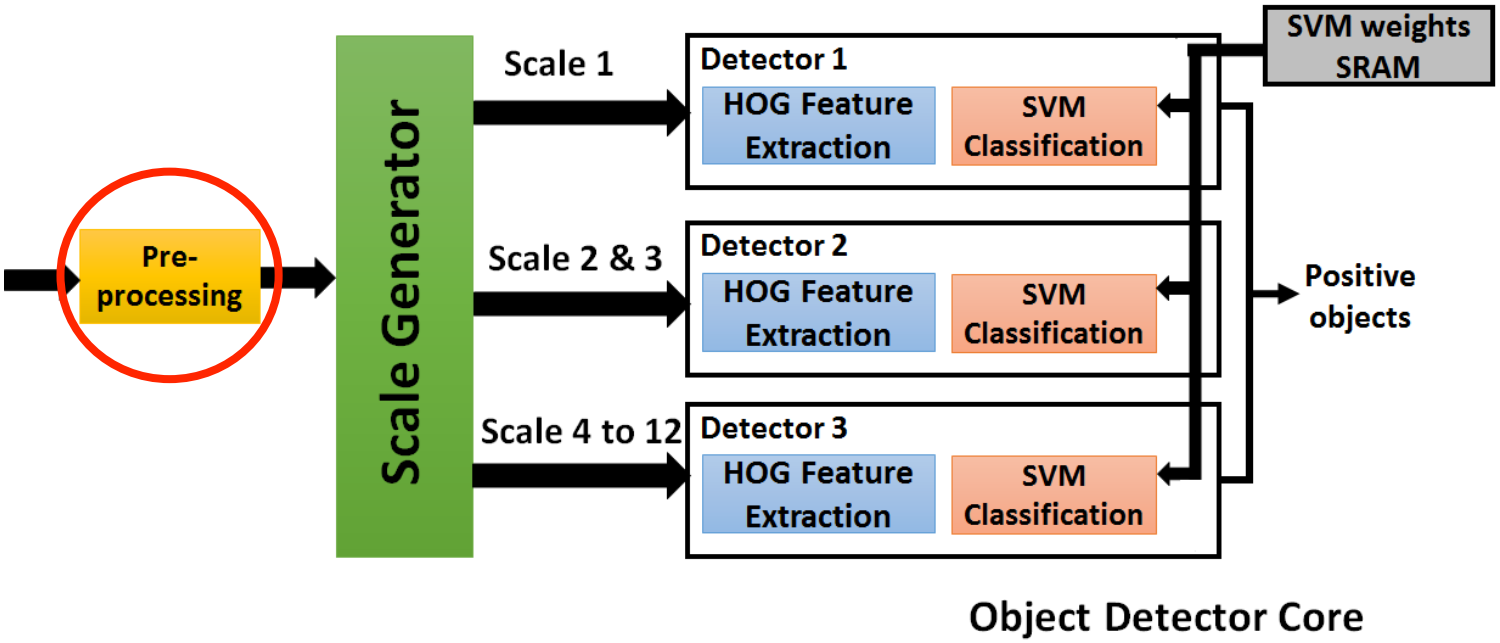


Image Pre-Processing

- Pre-process to reduce scales generation overhead
- Alternative representation with smaller pixel bit-width:
  - Smaller pixel buffer sizes
  - Simpler interpolation logic

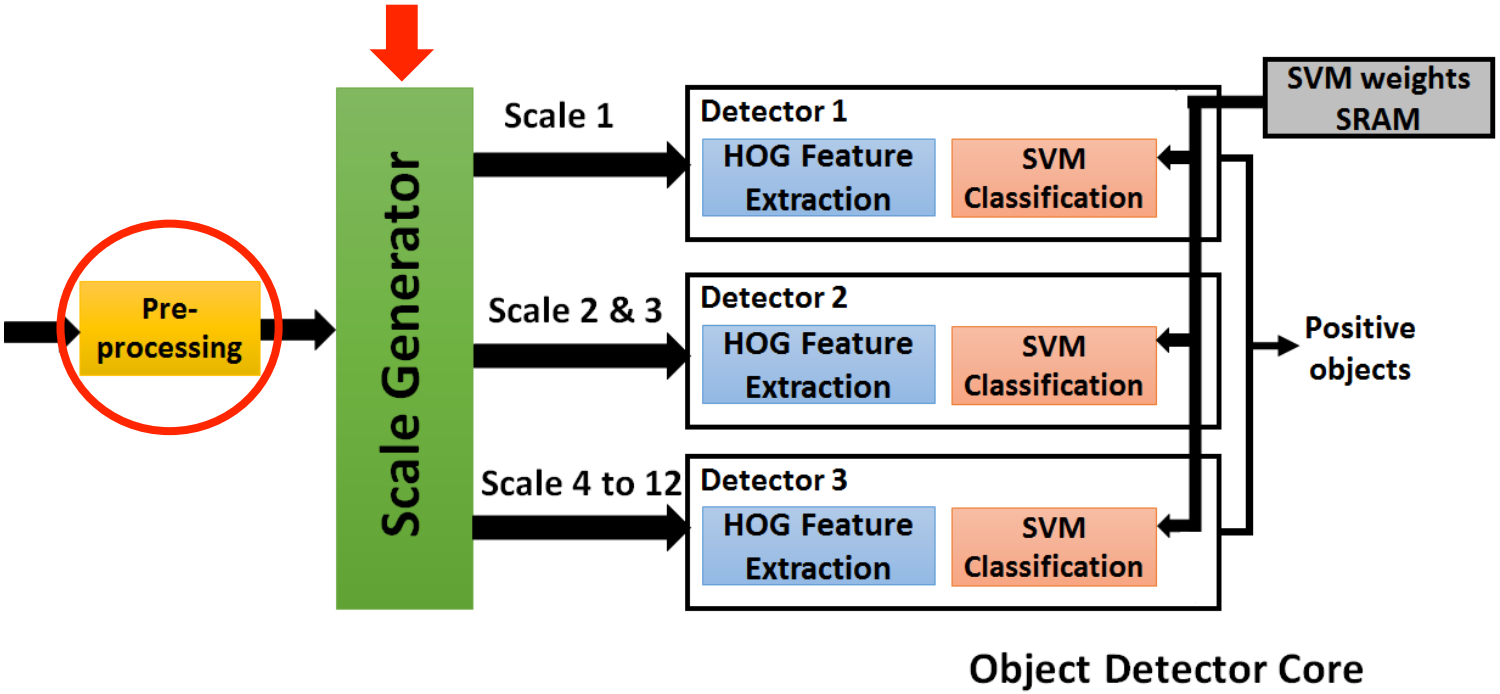


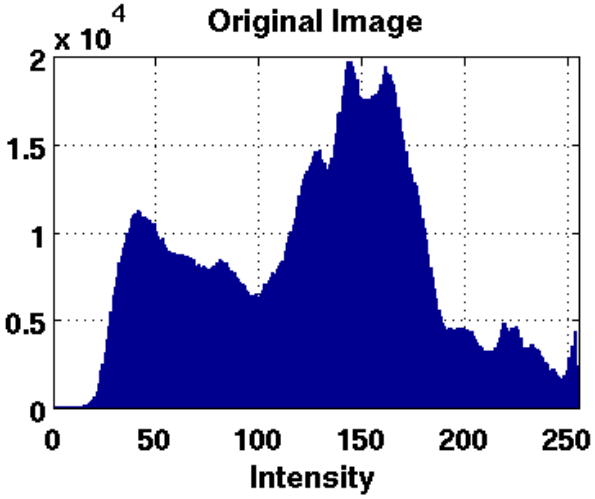
Image Pre-Processing

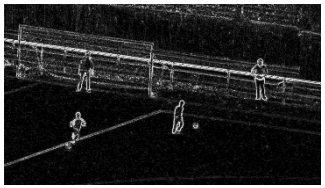
- Pre-process to reduce scales generation overhead
- Alternative representation with smaller pixel bit-width:
  - Smaller pixel buffer sizes
  - Simpler interpolation logic

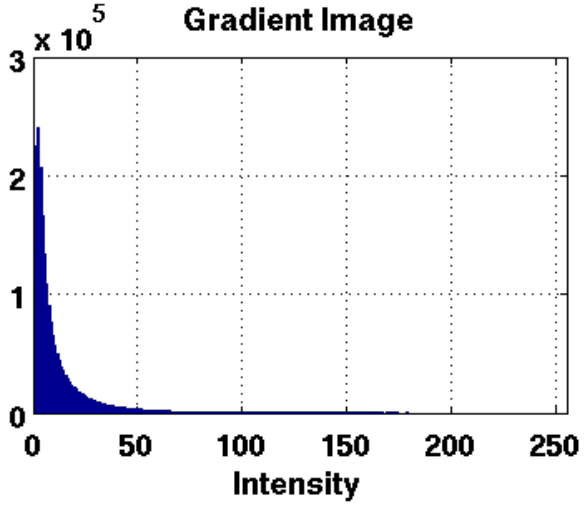
Original Image

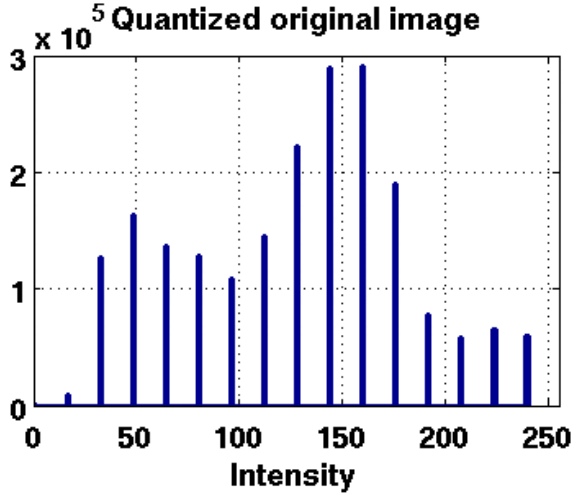
Gradient (edge) image

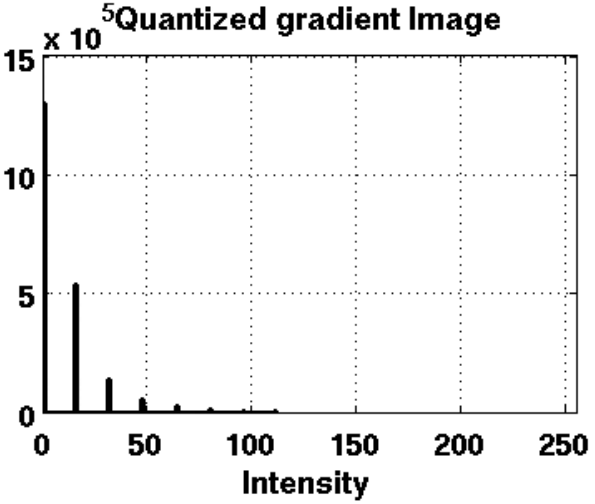
Image Pre-Processing





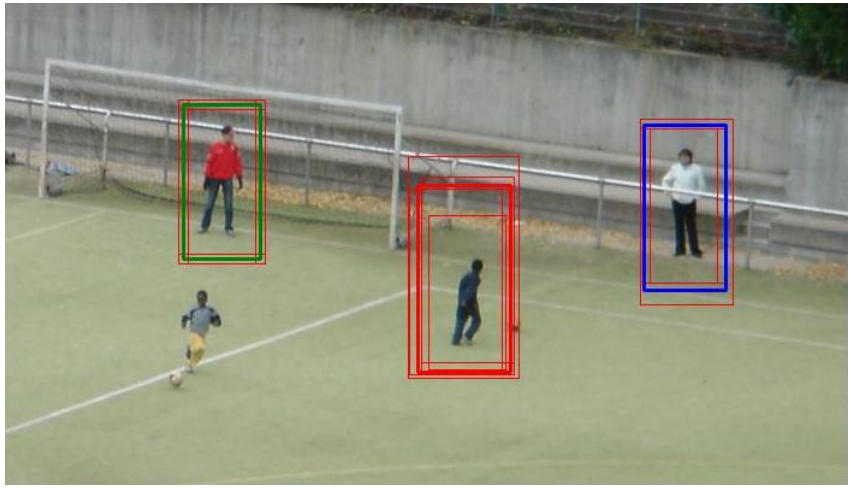


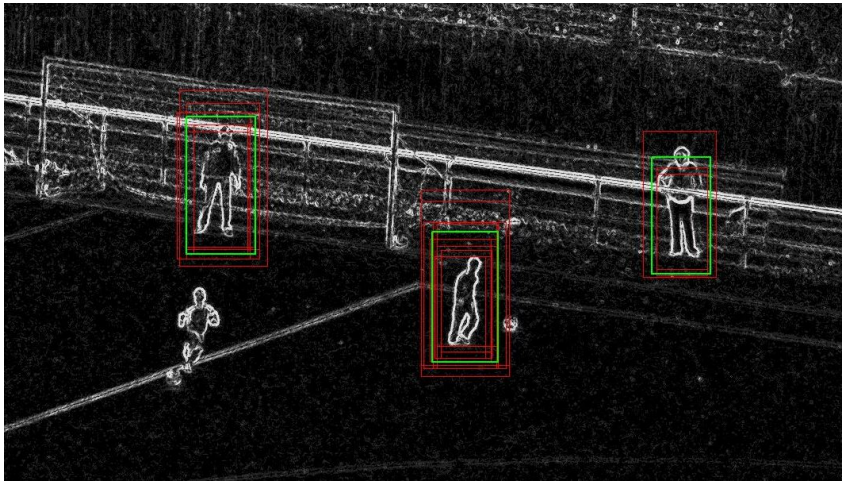




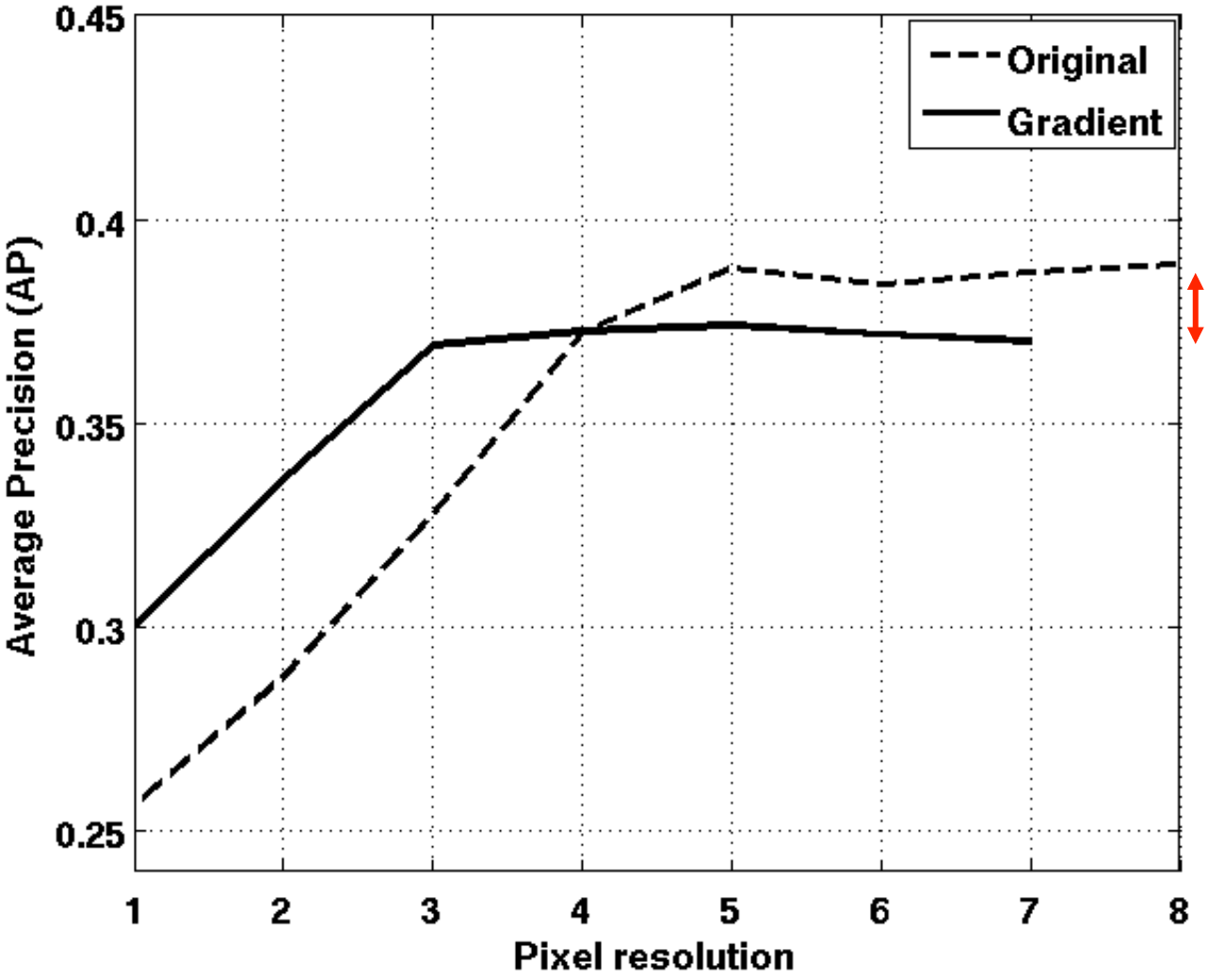
Detection on Gradient Images

- Edges orientation are still preserved in gradient images
- HOG features are similar



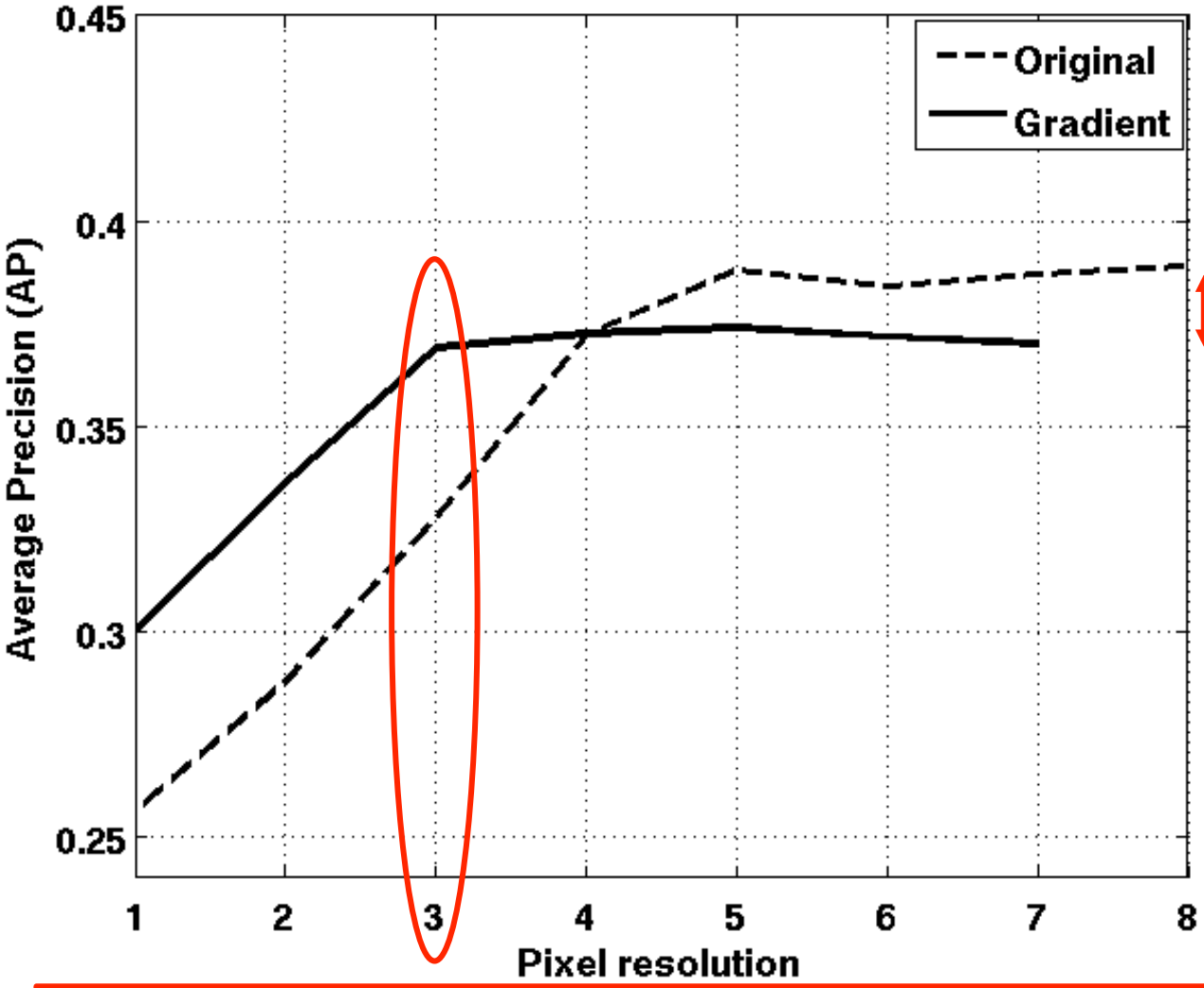


AP With Pre-Processing



Using INRIA person dataset

AP With Pre-Processing

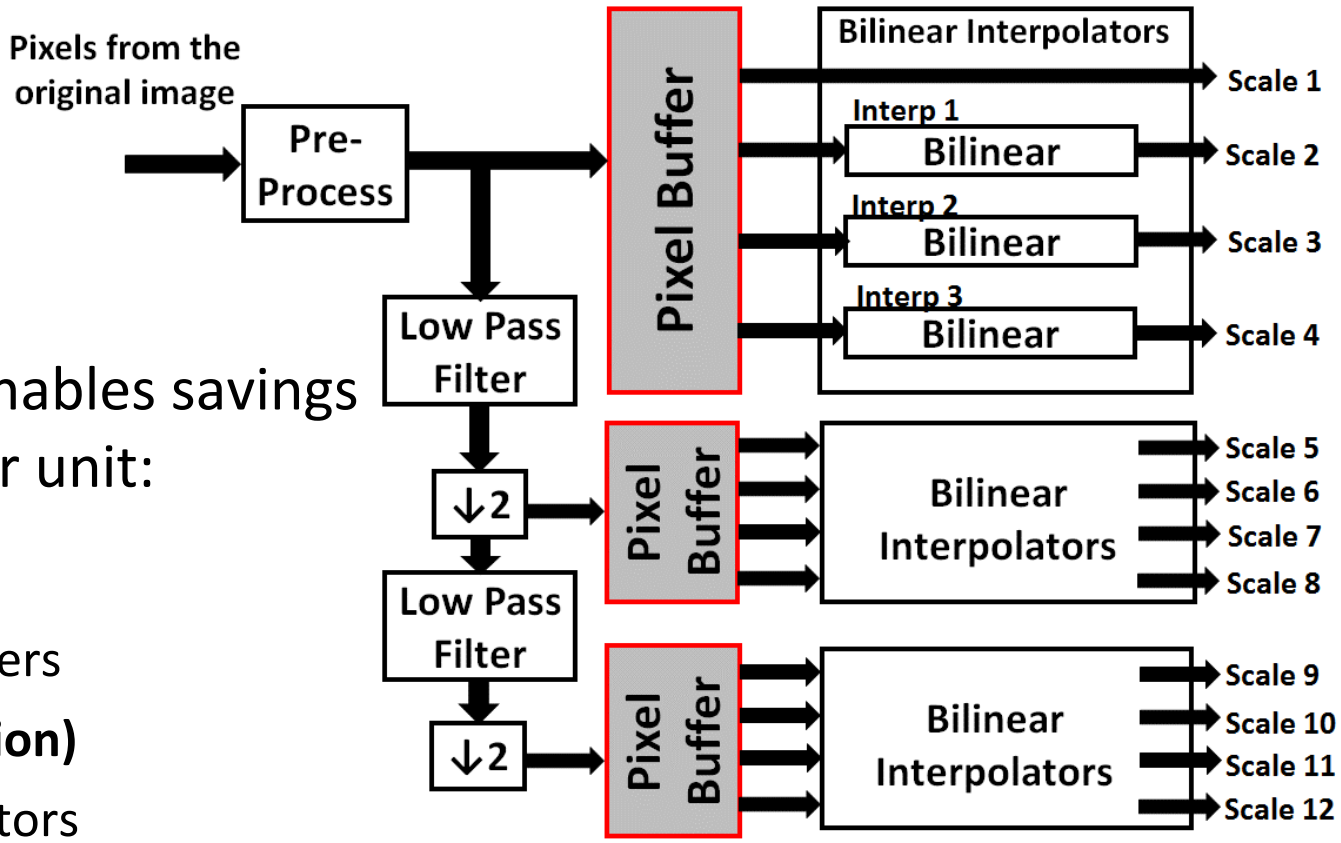


Using INRIA person dataset

Use gradient images with 3-bit pixels

Impact of Pre-Processing on Scale Generator

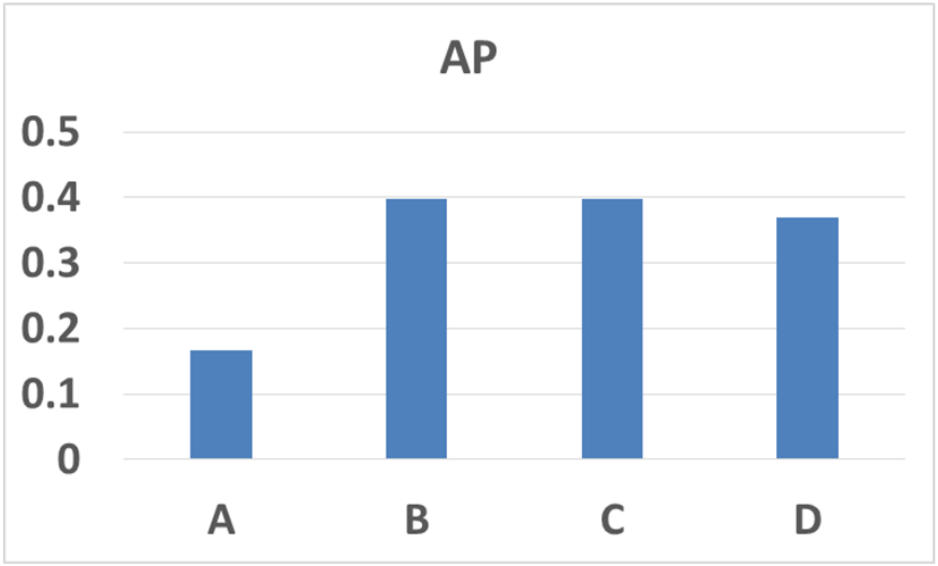
- Pre-processing enables savings in scale generator unit:
  - Smaller pixel buffers
    (2.67x reduction)
  - Simpler interpolators
    (43% power saving)

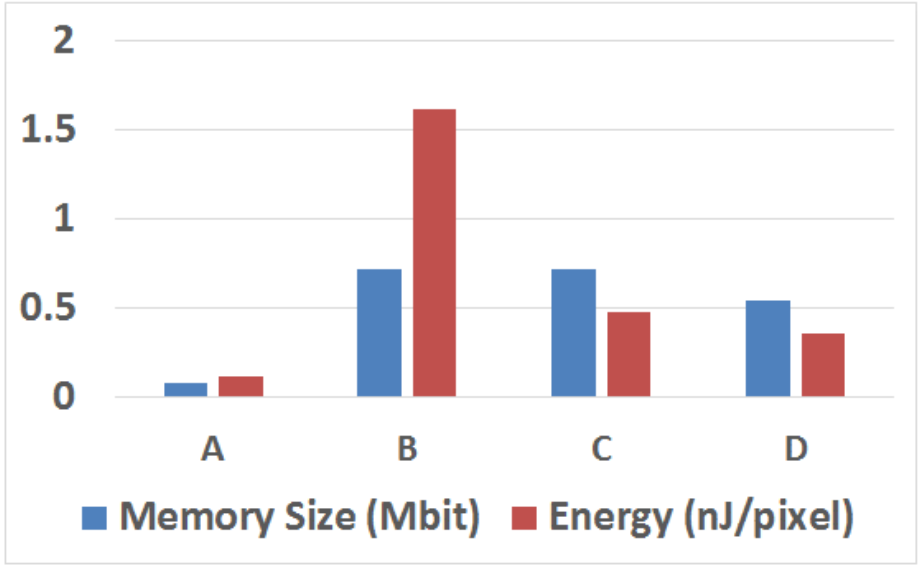


Scale Generator Unit

Pre-Processing Summary

- **Detection:** Reduced by **0.02 AP**
- **Energy:** Reduced by **24%**.
- **Memory:** Reduced by **25%**





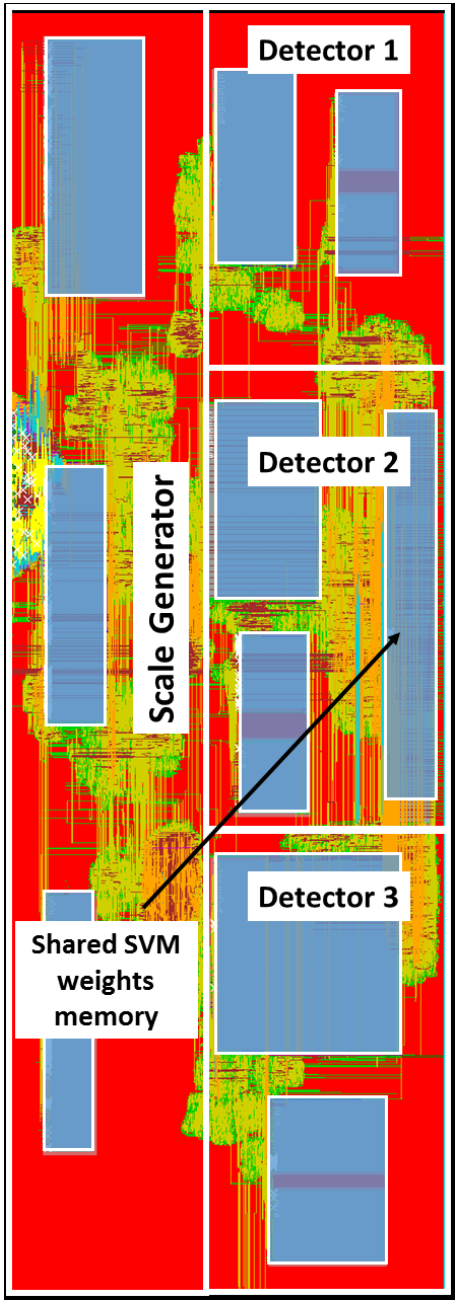
A: Single scale, single detector
B: Multi-scale, single detector
C: Multi-scale, 3-detectors
D: Multi-scale, 3-detectors, pre-processing

Constant throughput of 60 fps

Hardware Results
(Post-layout simulations)

- The circuit is designed in 45nm SOI CMOS

Area	2.80x0.96 mm^2
Max Frequency	270 MHz
Resolution	1920x1080 (1080HD)
Frame rate	60 fps
Scales/frame	12
Gate count	490 kgates
On-chip memory	0.538 Mbit
Power	45.3mW @ 0.72V
Energy	0.364 nJ/pixel



Conclusions

- An energy-efficient object detector is implemented delivering **real-time processing of 1080HD at 60 fps**
- Multi-scale support for **2.4x higher detection accuracy**
- Parallel detectors, voltage scaling and image pre-processing for **4.5x energy reduction.**