Deep Learning Visual Sensor for Industrial Applications

Jerry Byungik Ahn
Neurocoms Inc.
2018.7.21

Contents

About the Presenter

- Developer
- Computer architect
  - we value efficiency (speed/resource) rather than just speed
- Founder / CEO of Neurocoms Inc.


CPU

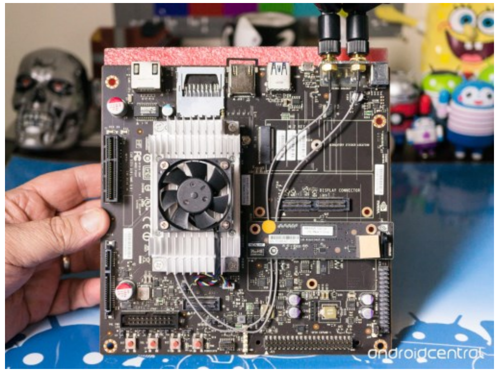

GPU


New Architecture

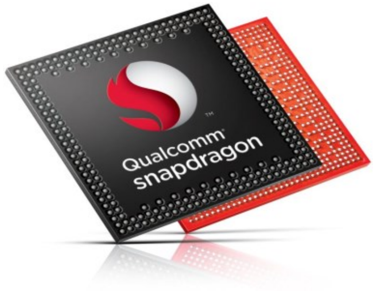
A Comparison of Mobile DL Hardware

NVIDIA GTX980M GPU

25.4 FPS

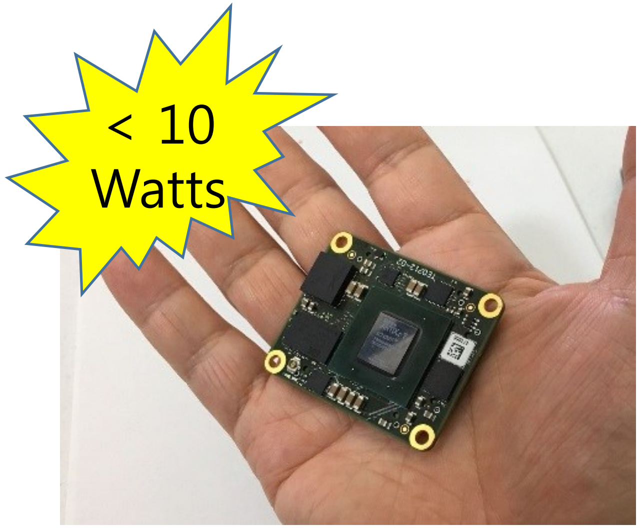

NVIDIA Jetson TX1 GPU

3.3 FPS



Qualcomm Snapdragon 820 with GPU

5 FPS

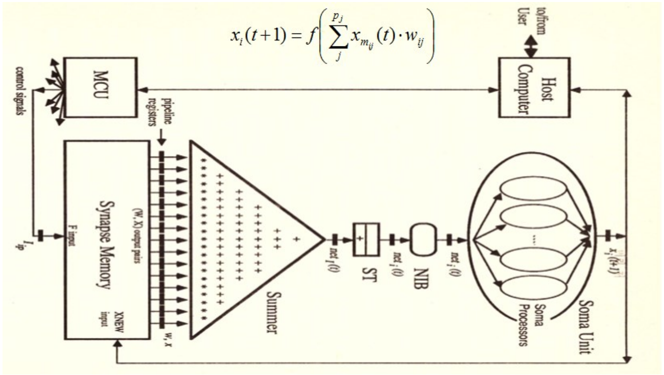


Neurocoms Deep Runner

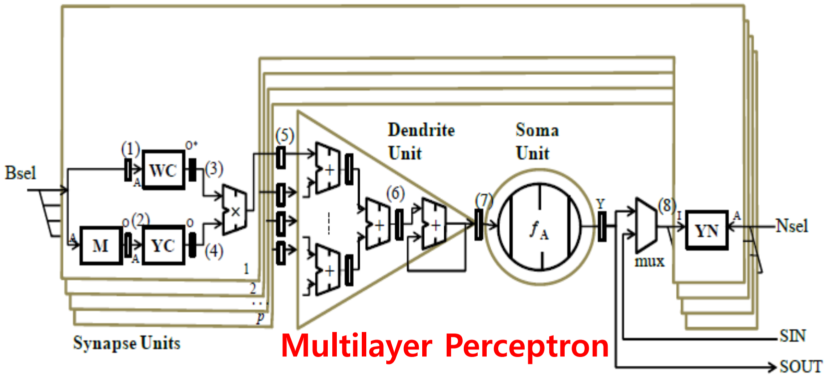
30 FPS

SSD300/MobileNet Object Detection

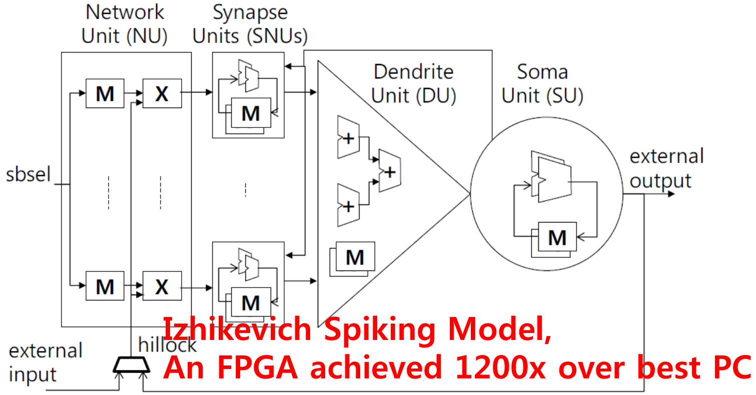
Hardware Architecture: Neuron Machine



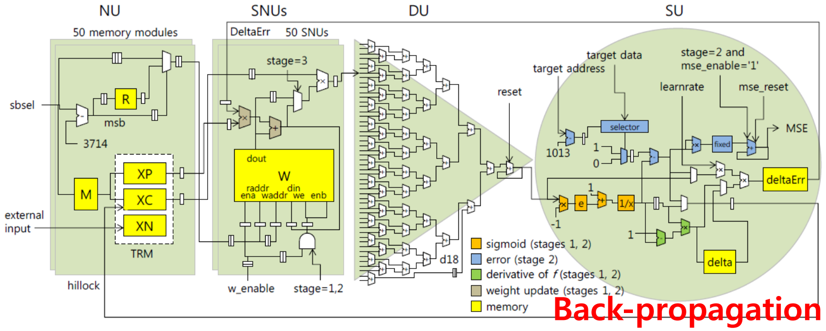
"*A study on a neuron model architecture for neurocomputing*", Master Thesis, KAIST, 1990



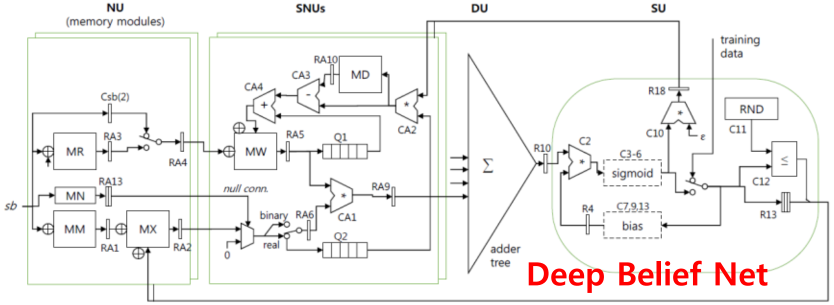
"*Neuron machine: Parallel and pipelined digital neurocomputing architecture*", IEEE Cybernetics Com, 2012



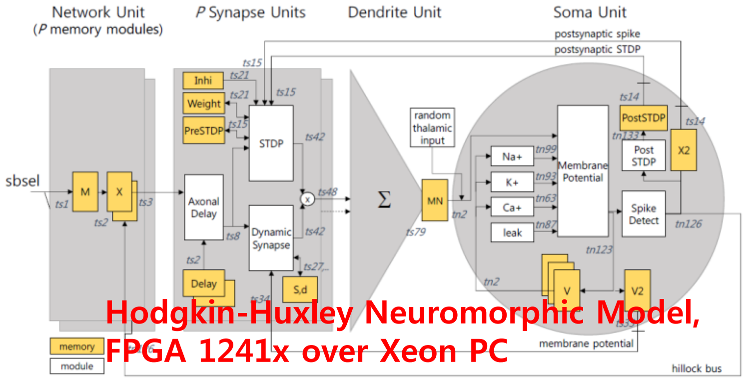
"*Extension of neuron machine neurocomputing architecture for spiking neural networks*", IEEE IJCNN, 2013



"*Computation of Backpropagation Learning Algorithm Using Neuron Machine Architecture*", IEEE IJCNN, 2013



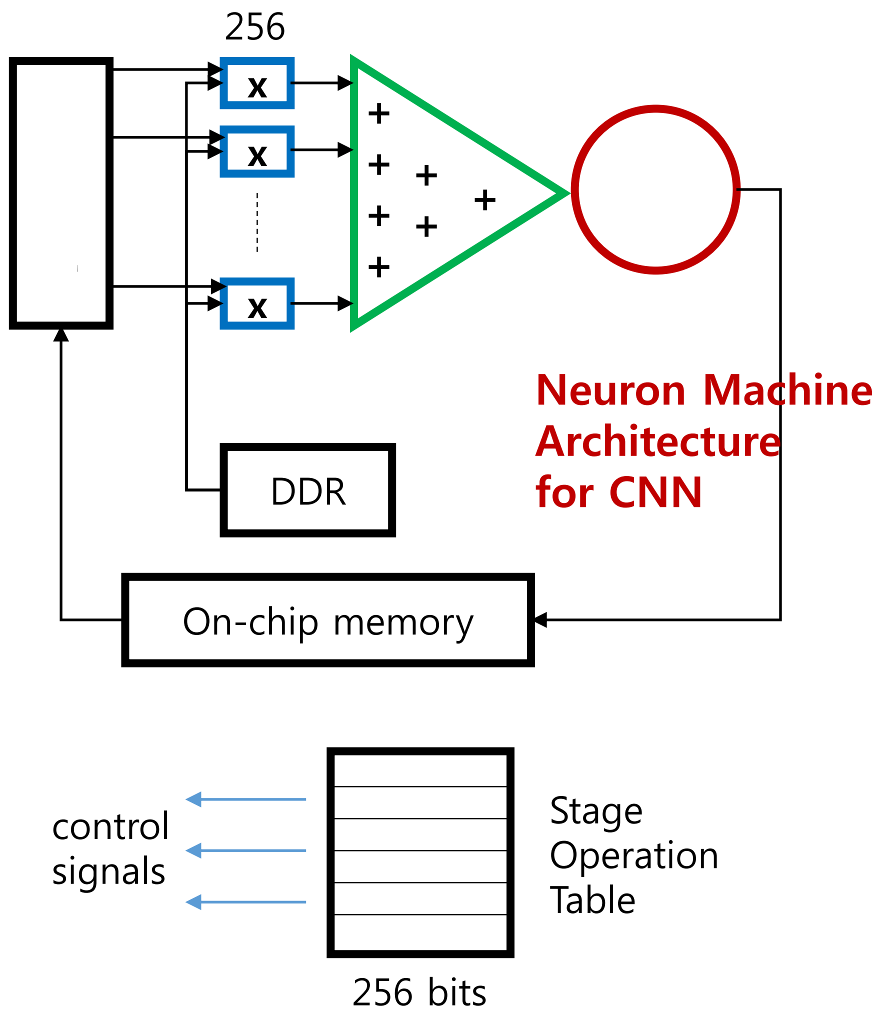
"*Computation of Deep Belief Networks Using Special-Purpose Hardware Architecture*", IEEE IJCNN, 2014



"*Neuron-like Digital Hardware Architecture for Large-scale Neuromorphic Computing*", IEEE IJCNN, 2015

Hardware Architecture: Neuron Machine

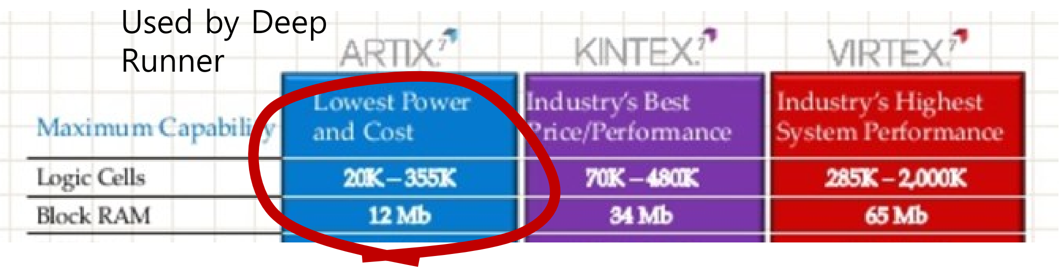
- Key ideas
  - Computational circuit same as the computation model (the shape of neuron)
  - Special memory circuit: (1) No inter-stage data movement required, (2) large number of slow memories
- Other properties
  - All self-contained in hardware: no processor involved
  - Fully pipelined and no idle clock cycle for arithmetic operators
  - High utilization of multipliers (see next page)
  - Sort of CISC computer - Each instruction for one CNN layer



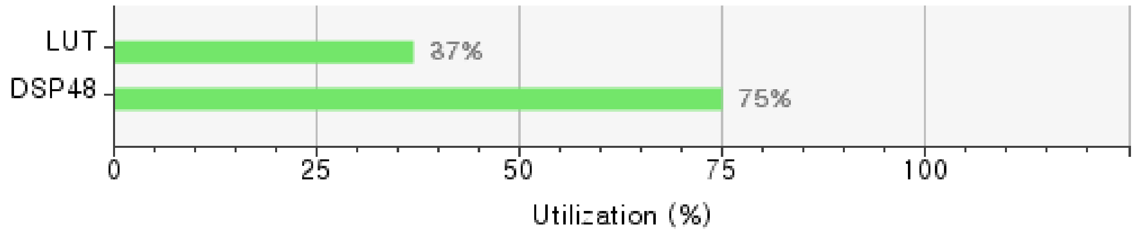
Multiplier Utilization Comparison

	A	B	C	D	E	F
Hardware	# of cores (multipliers)	Clock freq. (GHz)	Peak speed (AxB, Gops)	Optimal FPS (C/1.27[1)])	Actual FPS	Multiplier utilization (E/D)
GTX980M	1536	1.038	1594	1255	25.5	**2.03 %**
Jetson TX1	256	1.68	430	338.7	3.3	**0.96 %**
Deep Runner	256	0.2	51.2	40.3	29.5	**73.20 %**

1) 1.27 Giga operations are required for a single SSD300/MobileNet inference



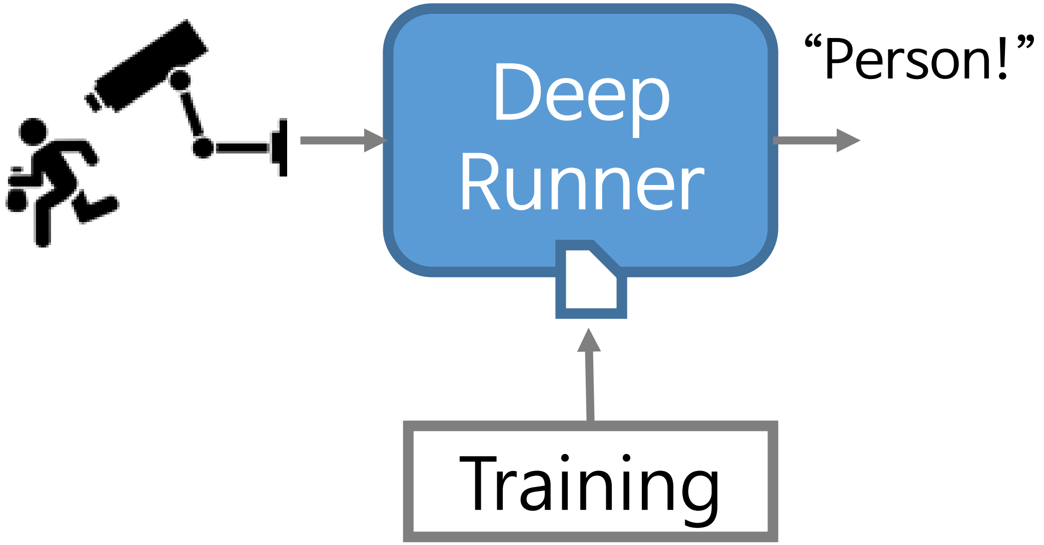
Xilinx 7-Series FPGA



Resource Utilization

Deep Runner Visual Sensor device

- Industrial device with built-in deep learning algorithms
  - To be used as a component for building intelligent systems
  - No prior knowledge of deep learning is required for user
- Supports multiple DL algorithms
  - GoogLeNet (classification)
  - YOLO, Tiny YOLO (detection)
  - SSD/MobileNet (detection)
  - MobileNet, Xception (classification)



Deep Runner Products

Deep Runner Module

- 4 x 5cm size
- Mounted as a part on user's PCB
- Input video signal: YCbCr4:2:2
- Recognition result: Ethernet, Serial port
- Power consumption: 5 watts

Deep Runner Device

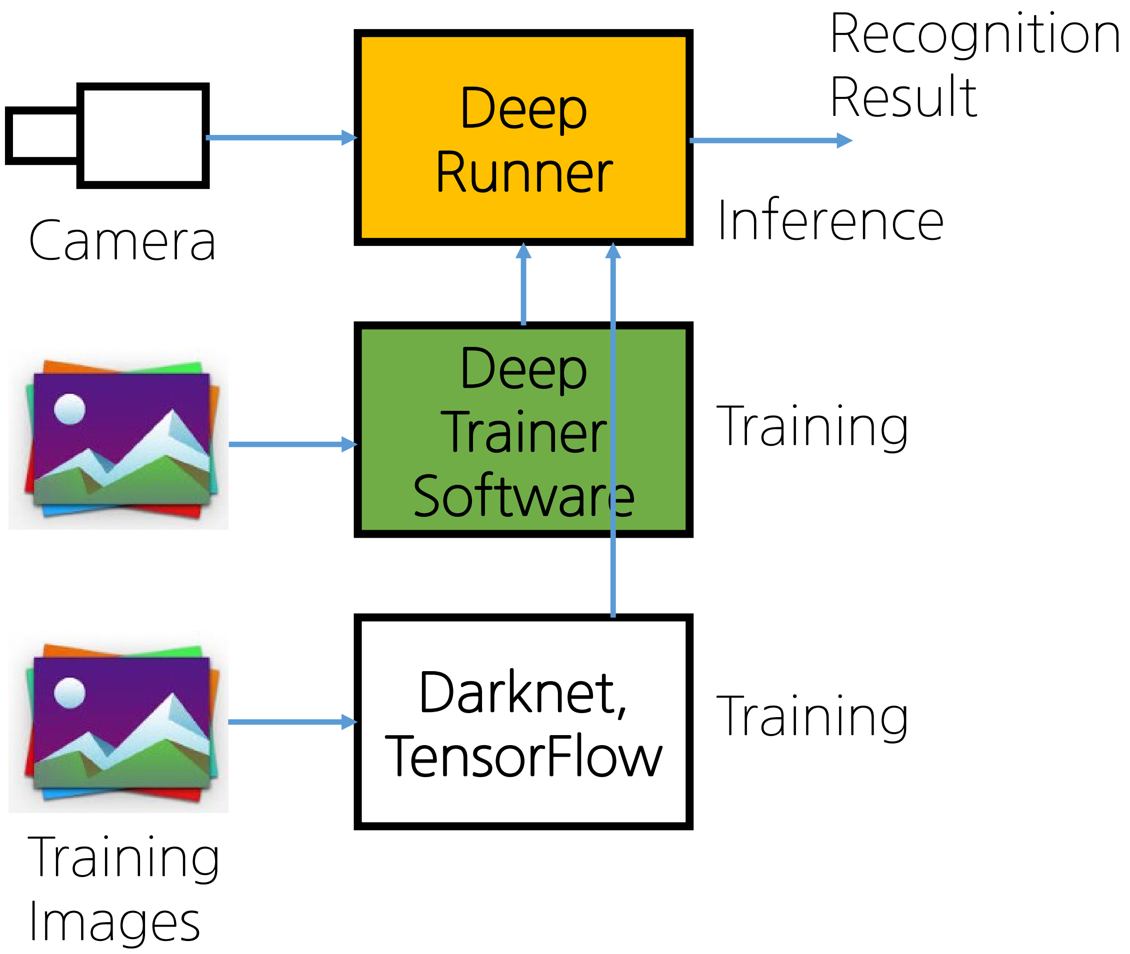
- Video Input: HDMI (1920x1080@30, 1600x900, 1280x720)
- Recognition Result: Ethernet, Serial port, GPIO pinout
- Simultaneous recognition of up to 16 cameras from split screen
- Power consumption: 8 watts

Deep Runner CCTV

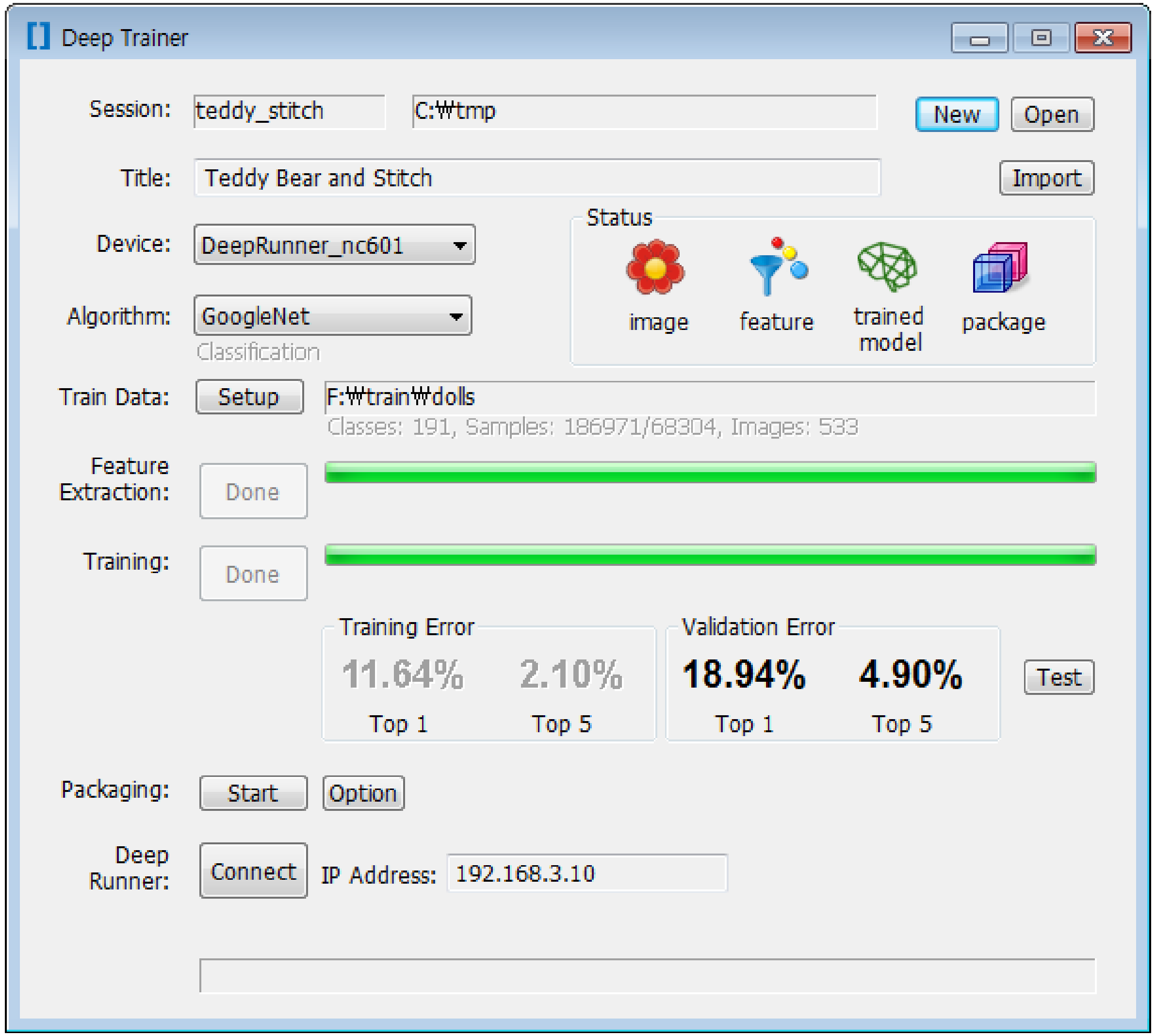
- For CCTV Surveillance
- Input: IP cameras
- Recognition Result: Ethernet
- Simultaneous recognition of up to 8 IP cameras
- Video recording function

Training Procedure

- Deep Trainer
  - Windows software
  - Train classification algorithms
- Darknet
  - Train YOLO and Tiny YOLO object detection algorithms
- TensorFlow
  - Train SSD object detection algorithm



Deep Trainer

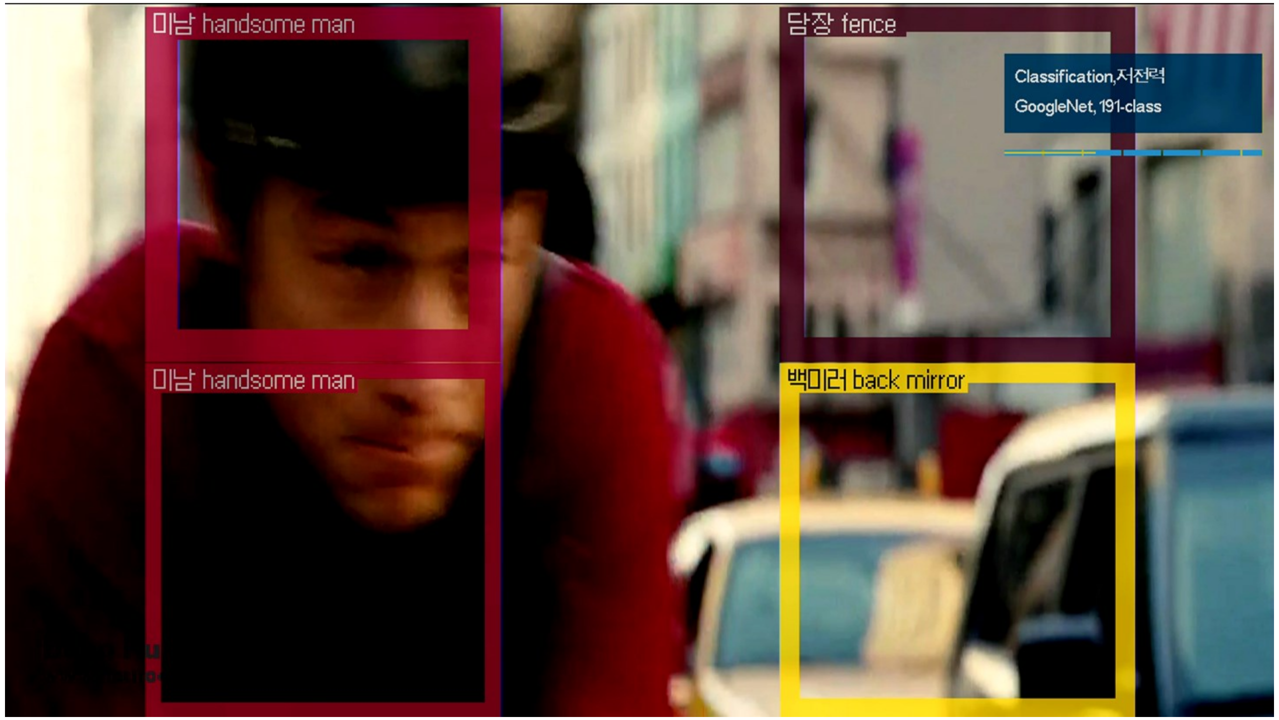


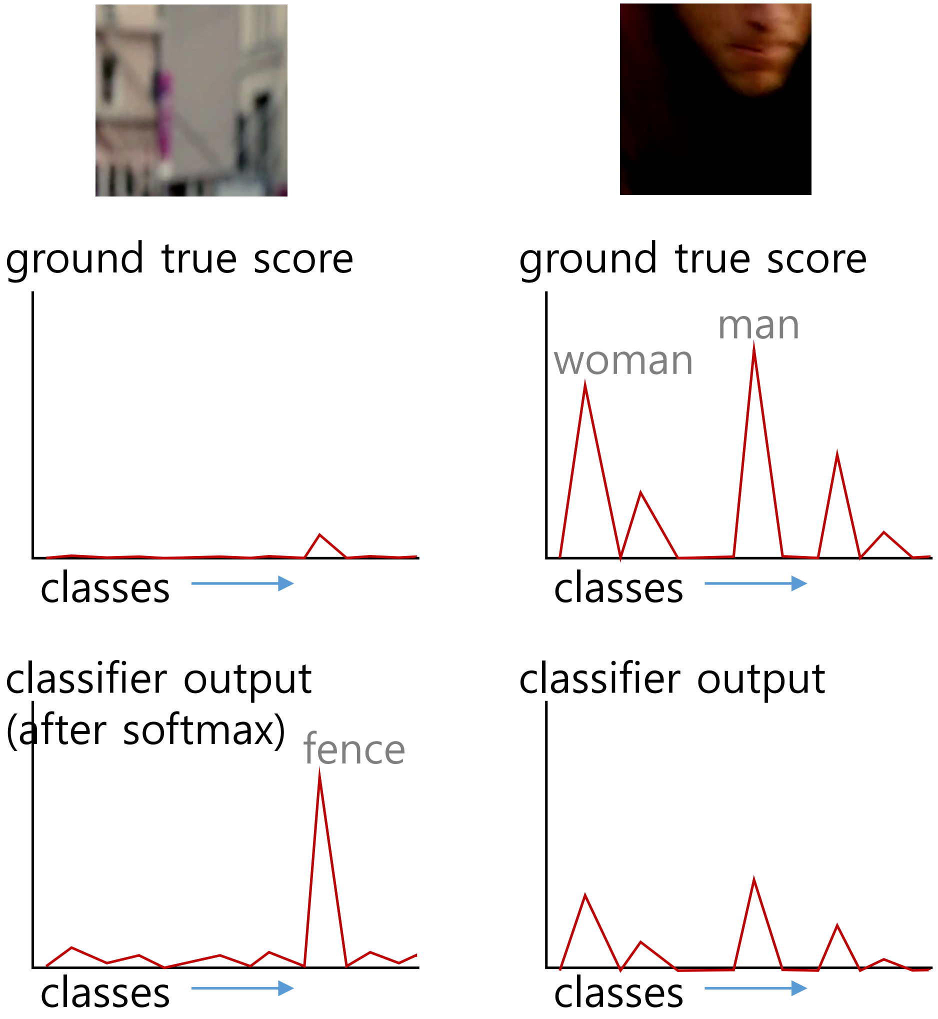
Reference Users

Customers	Area	Use
C	CCTV Surveillance	Show Room
N	CCTV Surveillance	Highway
S	CCTV Surveillance	CCTV Monitoring Center
S	CCTV Surveillance	
V, Turkey	CCTV Surveillance	
O, Japan	CCTV Surveillance	Security Camera
U	CCTV Surveillance	
I, Spain	CCTV Surveillance	Cloud system
H, Poland	Quality Inspection	
W, China	Quality Inspection	
T	Quality Inspection	
S	Automotive	Digital Room Mirror
H	Automotive	Around view of Excavators
H	Automotive	Around view of Excavators
E, A Univ.,S Univ.	Research, Education	

Limitation of the use of classification

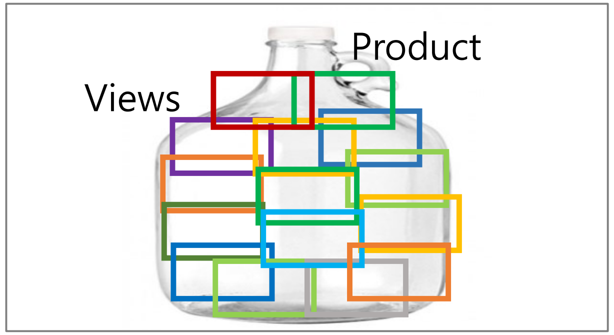
Because of the softmax function, the output of the classifier does not indicates exact score.



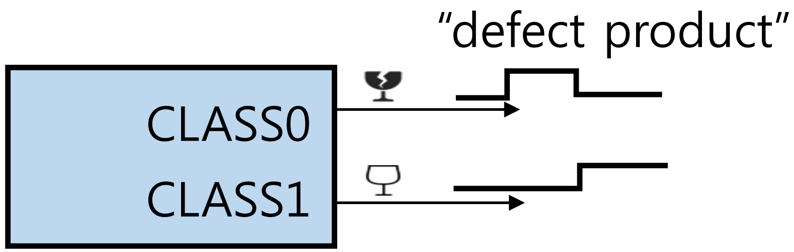


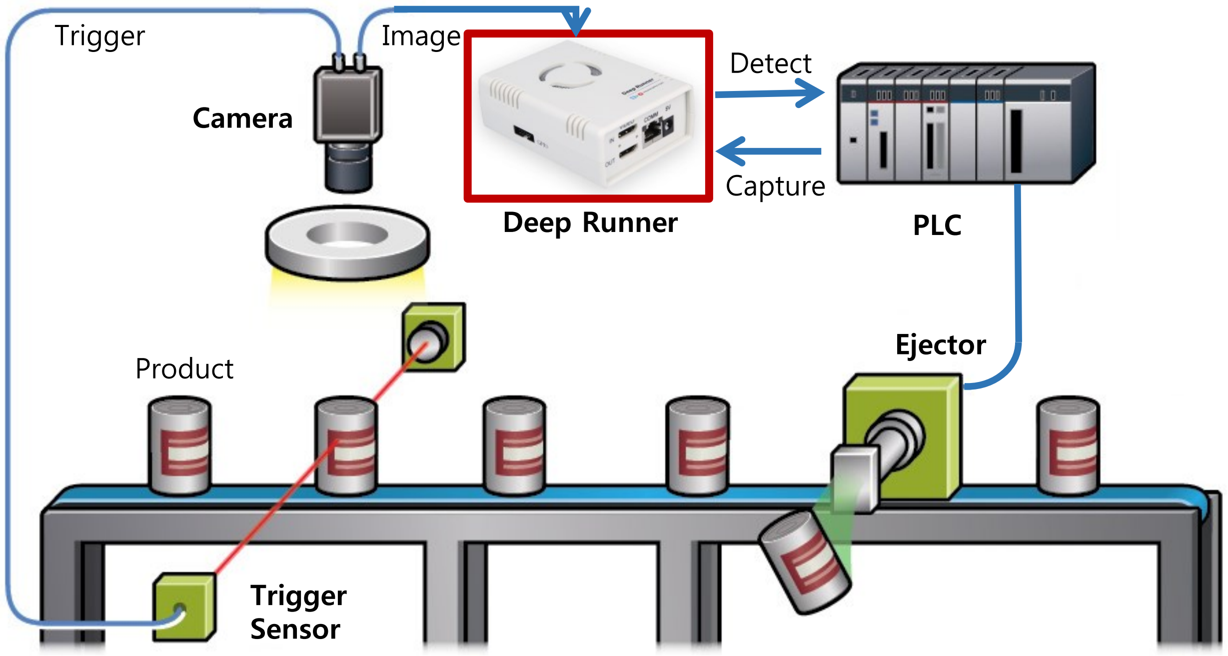
Quality Inspection

- The use of classification
- Special features
  - Find small defects in high resolution product images



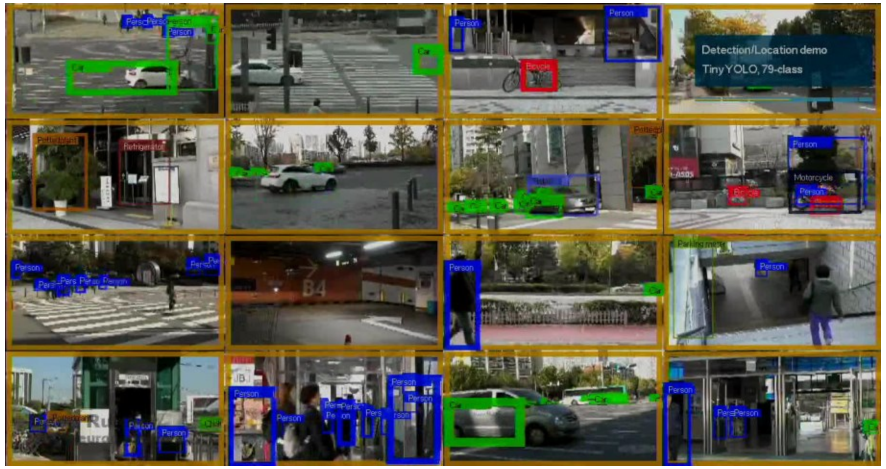
  - PLC pinout communication

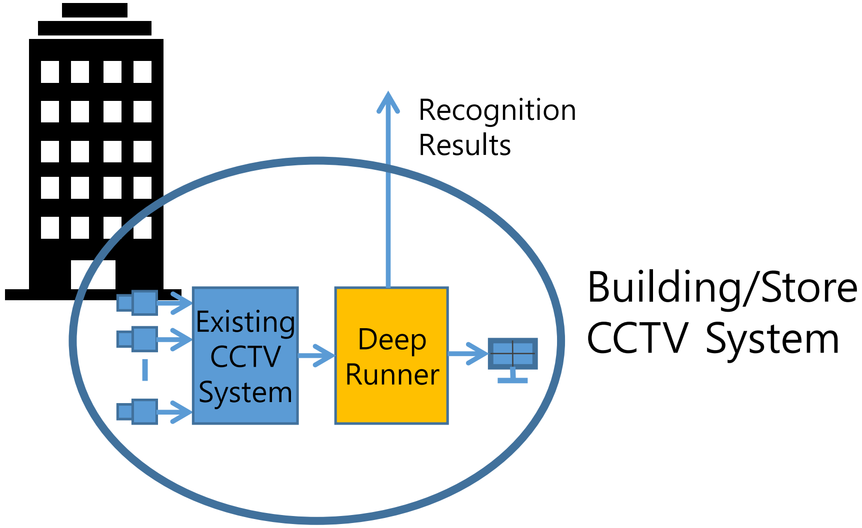


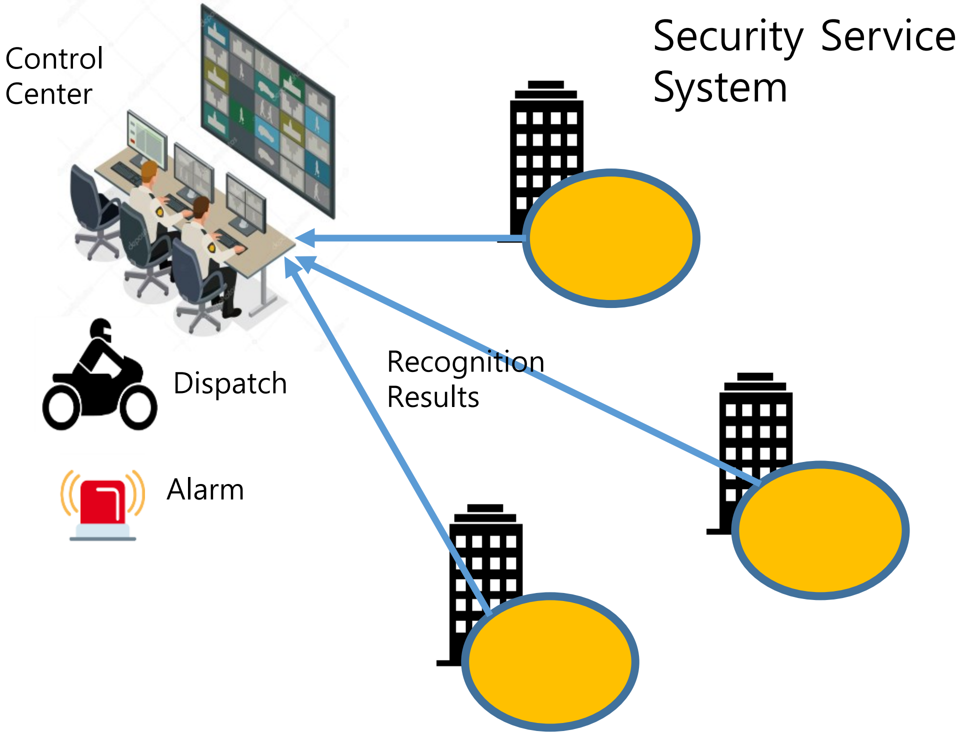


CCTV Surveillance

- The use of object detection
- Special feature
  - Recognize multiple cameras simultaneously







Conclusion

- Embedded deep learning will become mainstream
- As a leading company, we shared
  - Our hardware architecture
  - Device specification
  - Applications
- We are seeking
  - Funding for ASIC
    - Typically 50 times more power efficiency could be achieved with ASIC - 200mW with the same speed as Deep Runner
    - Applicable to millions of CCTV equipment
  - Collaboration projects
  - Recruits
  - Distributors