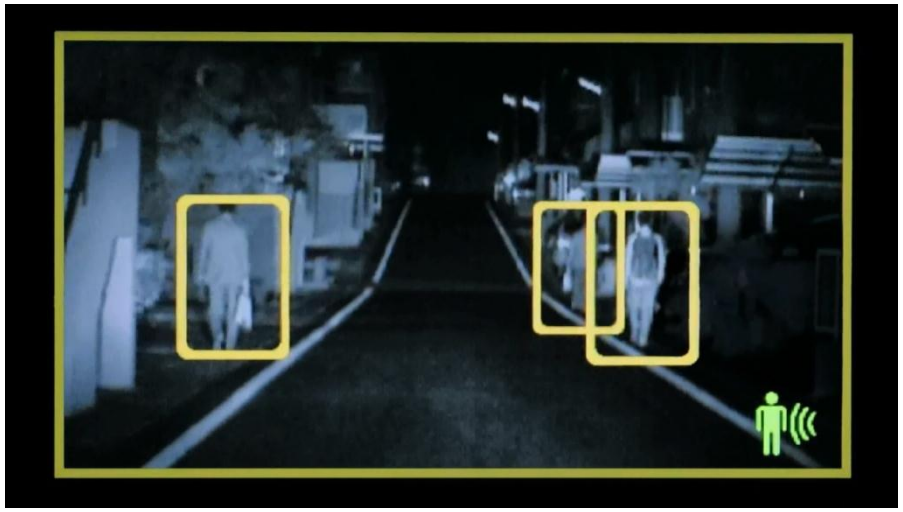
Beyond Pedestrian Detection: Deep Neural Networks Level-Up Automotive Safety

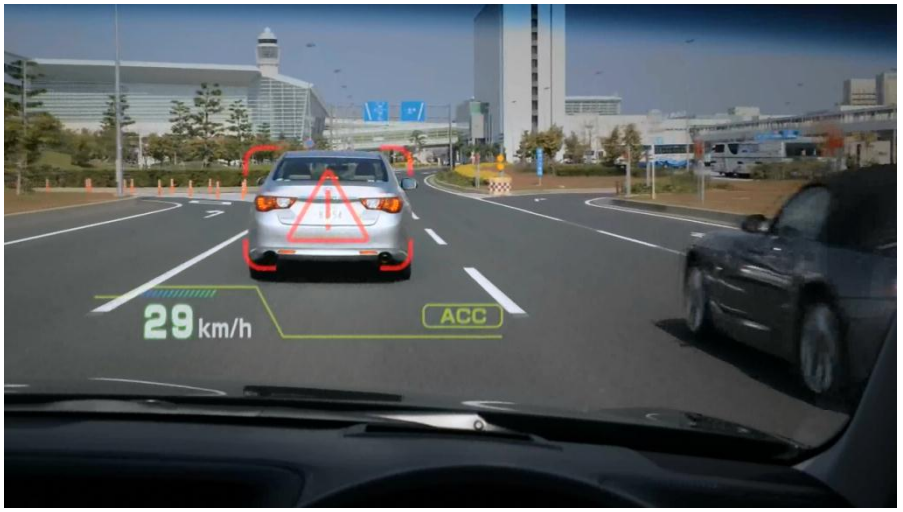
Ikuro Sato, Hideki Niihara

R&D Group, Denso IT Laboratory, Inc.

Advanced Driver Assistance Systems (ADAS)



pedestrian detection with infrared camera

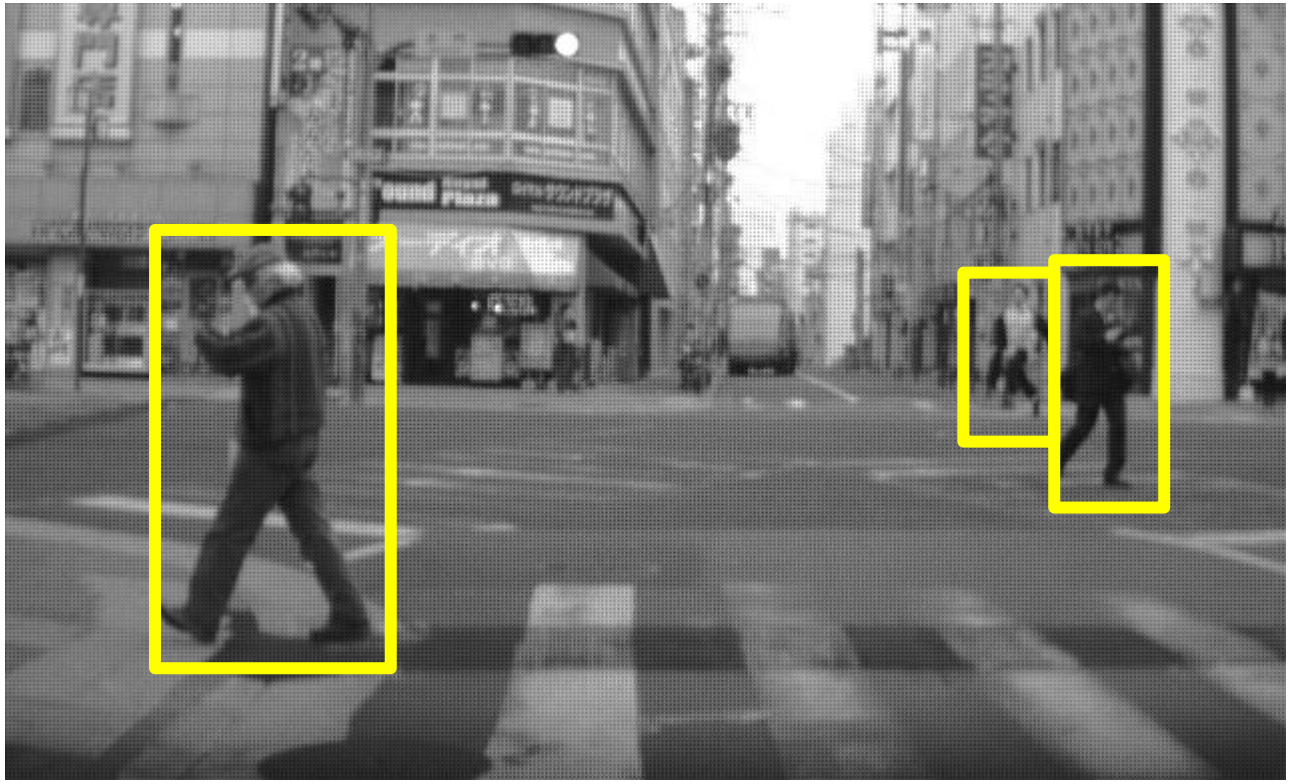


forerunning car detection with radar

Images taken from DENSO's TV ad.

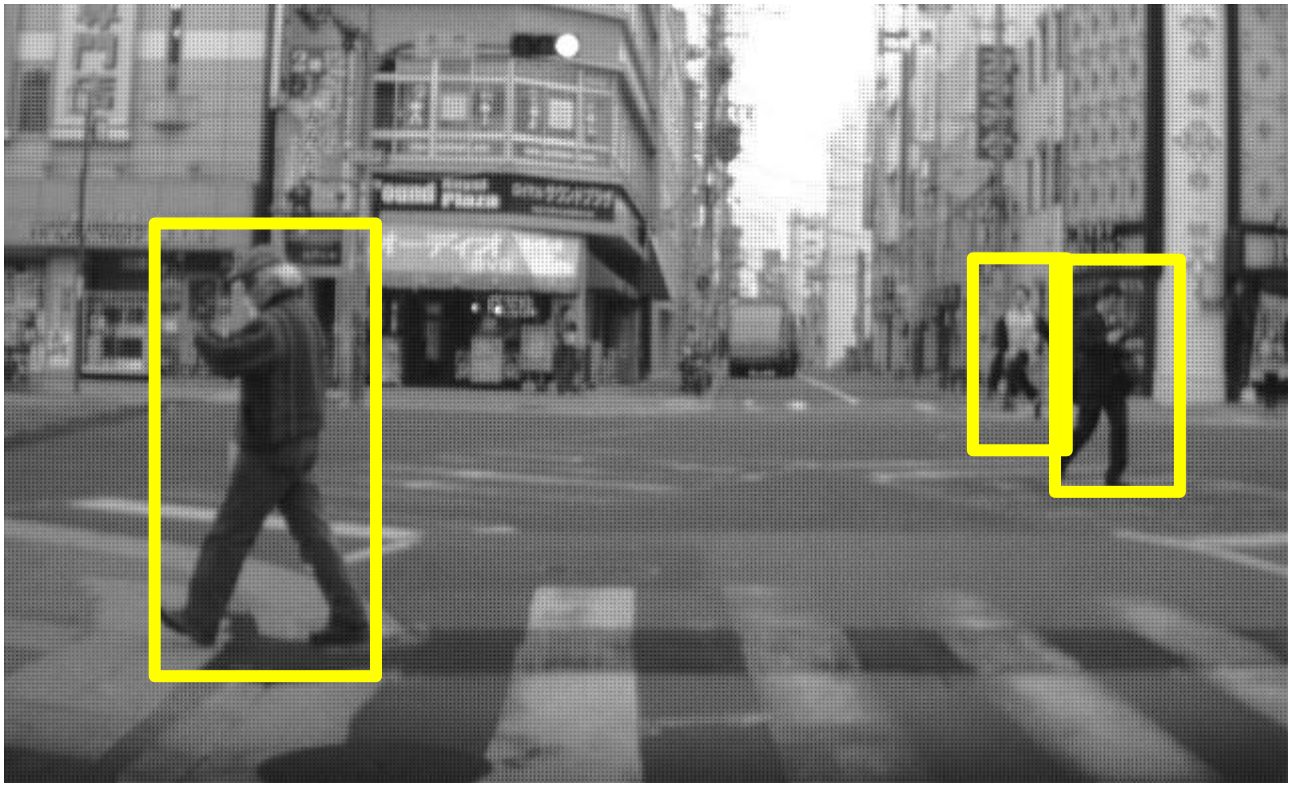
What would a skilled driver do?

What would a skilled driver do?



Today's Technology

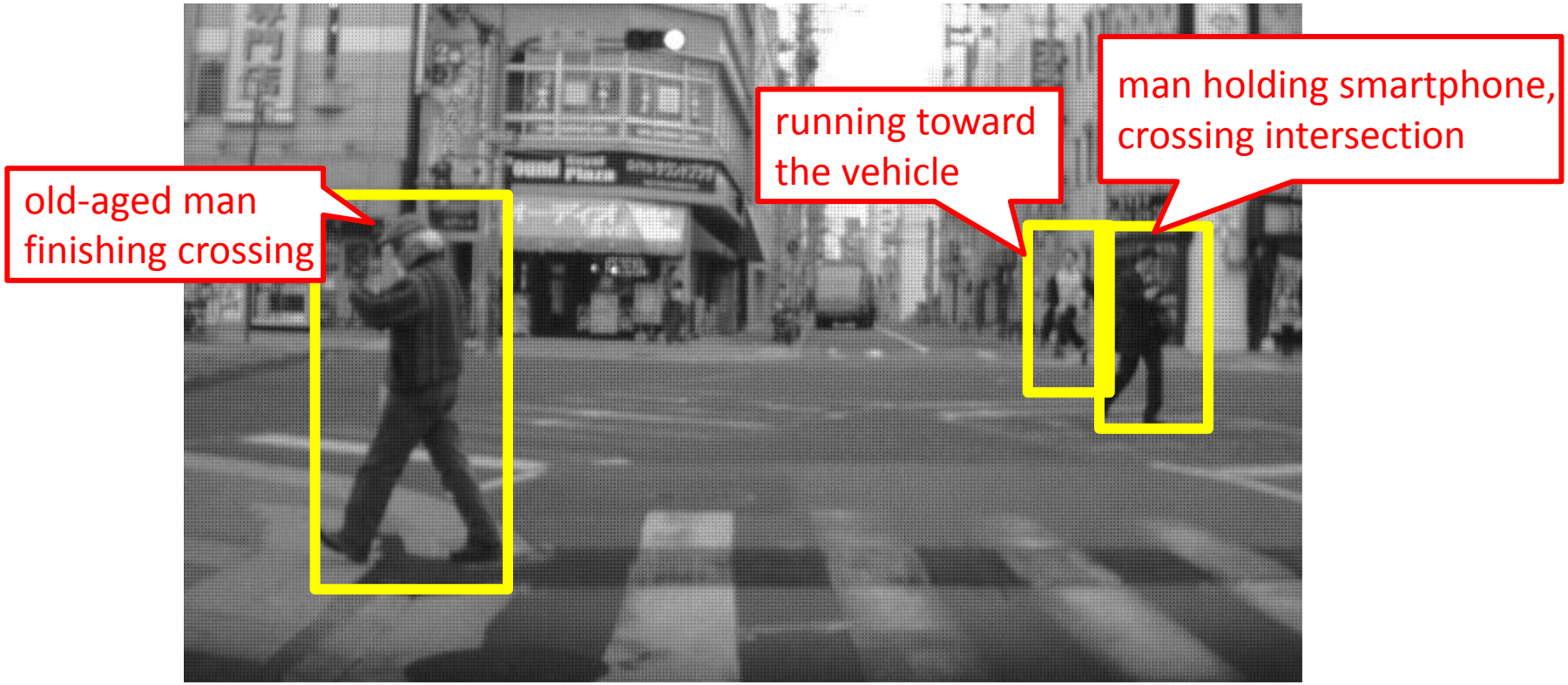
- Only to *detect* pedestrians with a camera
  - limited autonomous maneuvers possible, e.g., slows down to stop



today's industrial technology

The Goal

- Able to *extract richer information* about pedestrians
  - various autonomous maneuvers possible like a skilled driver



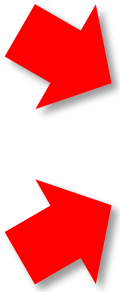
future industrial technology

Three Technical Challenges

- **multiclass classification**
  - Need to identify various annotative info on top of binary classification
- **visual feature design**
  - Who knows what visual features best separate a man with a smartphone from a man without?
- **real-time processing**

Promising Tools

- **multiclass classification**
- **visual feature design**



Deep Neural Networks

- **real-time processing**

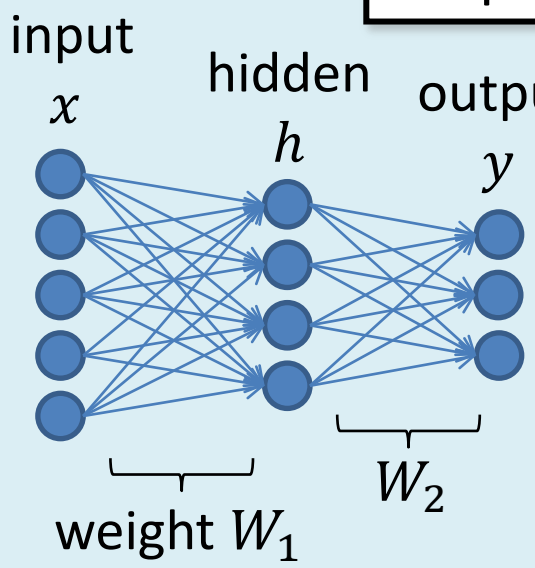
GPGPUs

- **Why Deep Neural Networks?**
- Why GPGPUs?
- Demo Setup

Basics

- **Neural Network**: one of the machine learning algorithms for classification and/or regression

simple Neural Network



$$\begin{cases} h = g(W_1 x) \\ y = g(W_2 h) \end{cases}$$

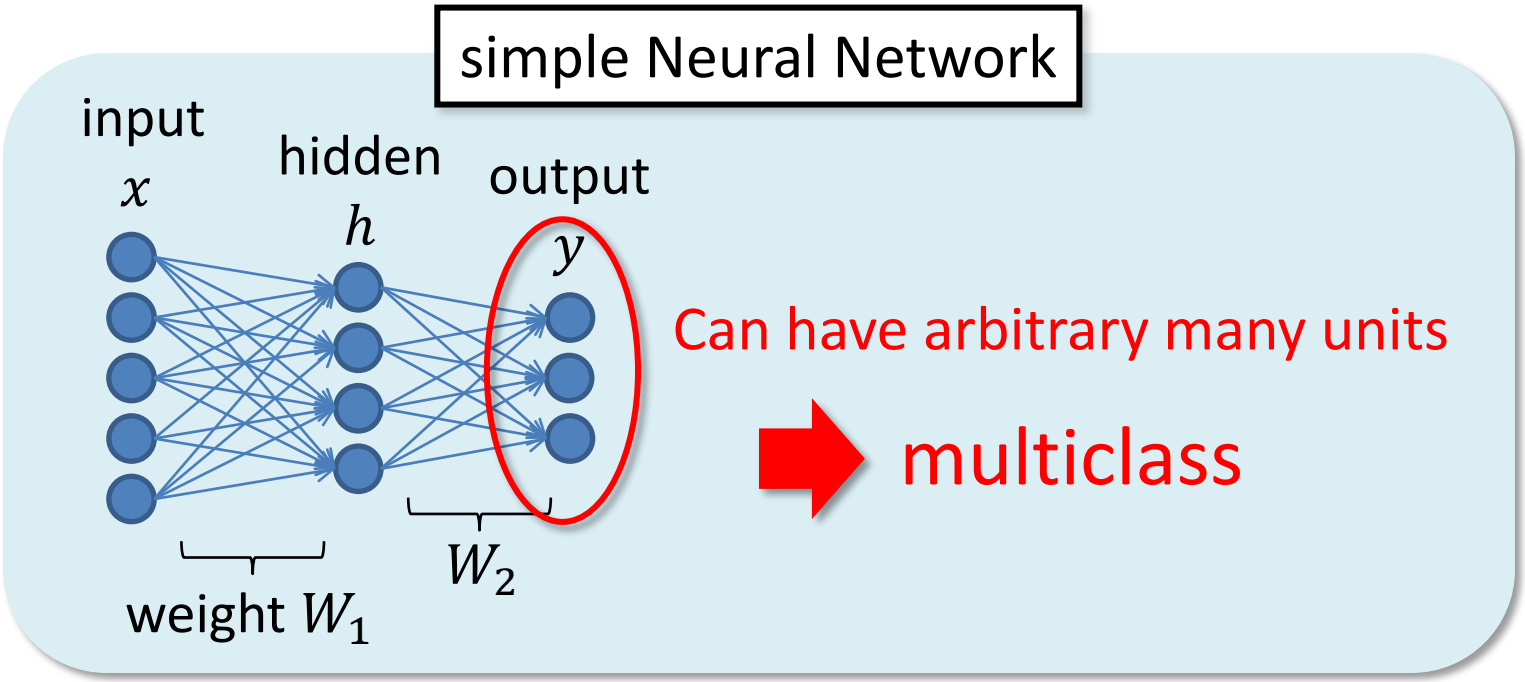
(simplified)

g: differentiable nonlinear function

By tuning W_k, one obtains an arbitrary function.

Basics

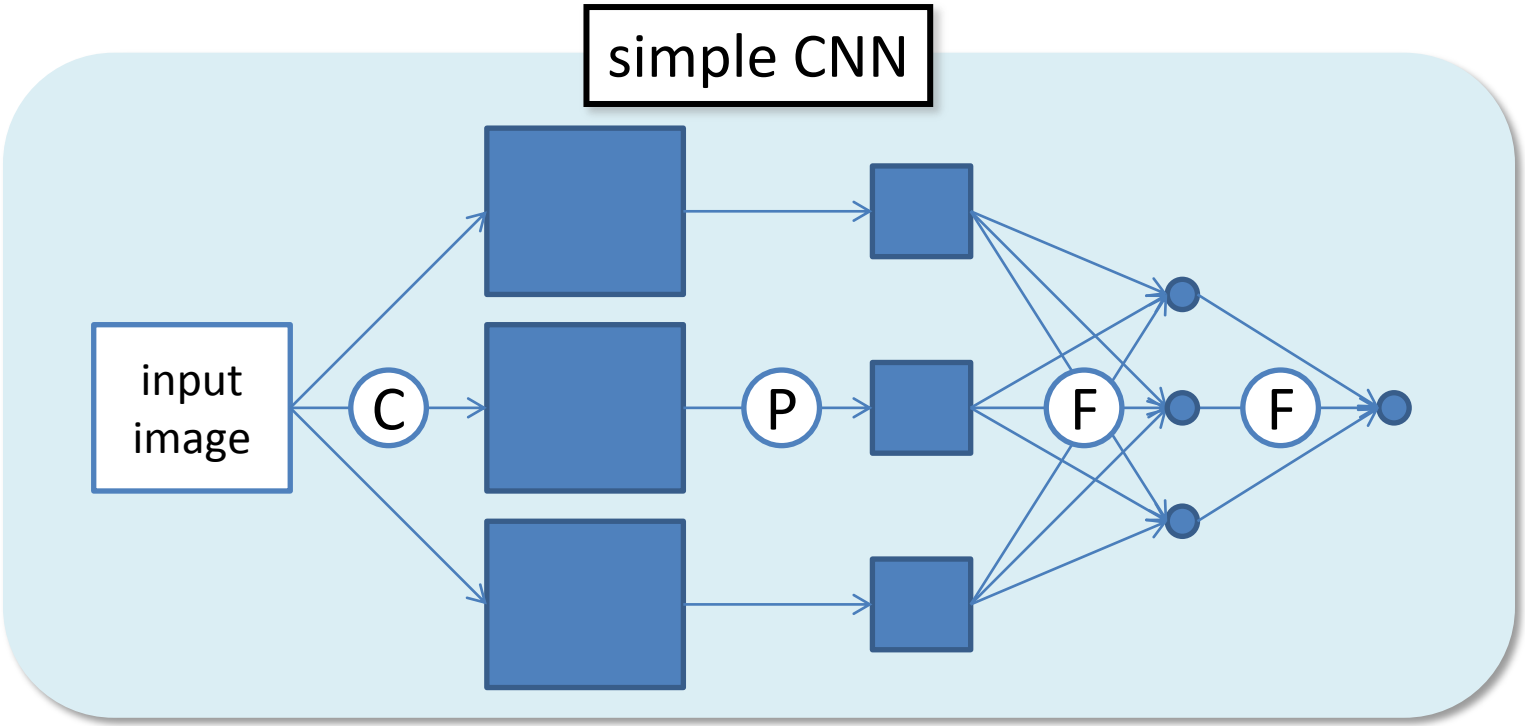
- **Neural Network**: one of the machine learning algorithms for classification and/or regression



By tuning W_k, one obtains an arbitrary function.

Neural Networks for Visual Tasks

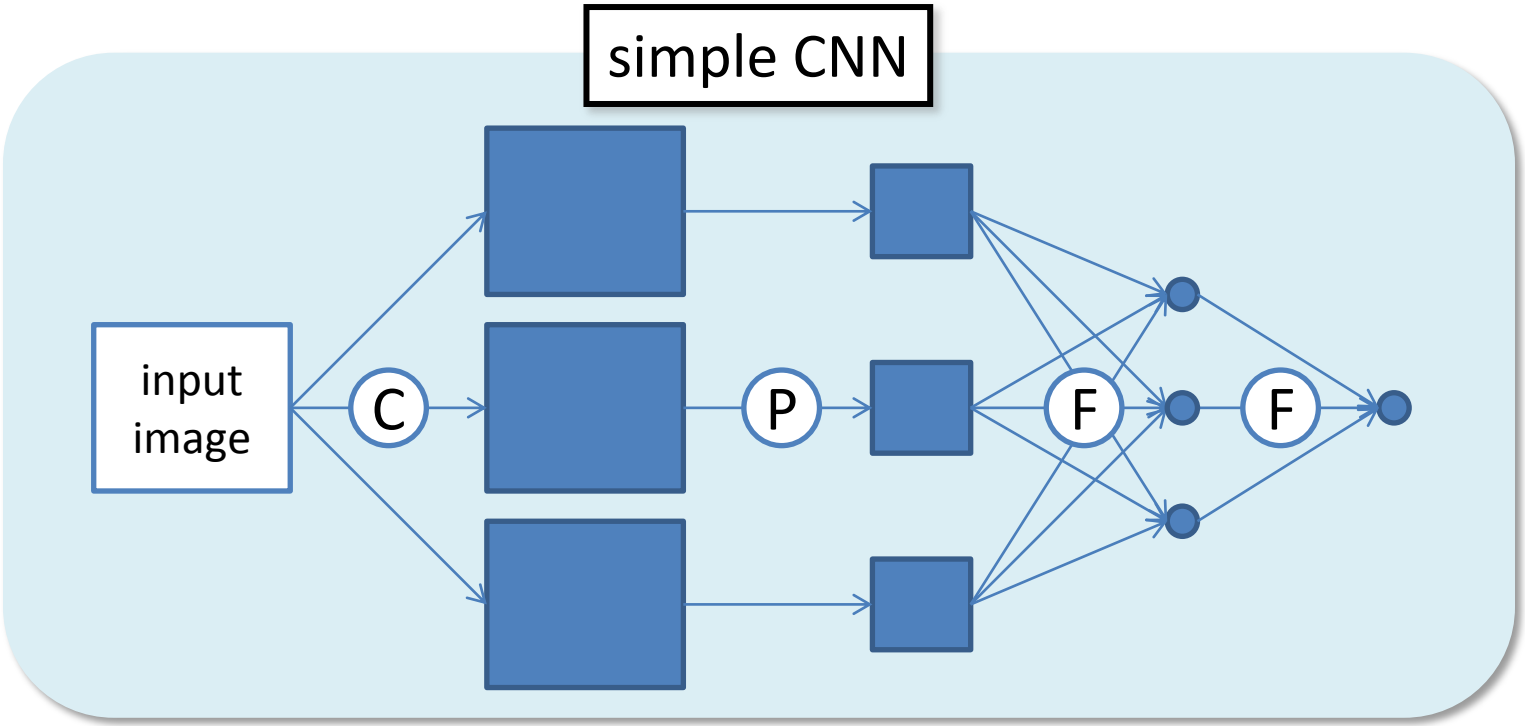
- **Convolutional Neural Network** (CNN) comprising
  - **Convolutional** layer(s) → local feature extraction
  - **Pooling** layer(s) → dimensionality reduction
  - **Fully-connected** layer(s) → classification/regression



Neural Networks for Visual Tasks

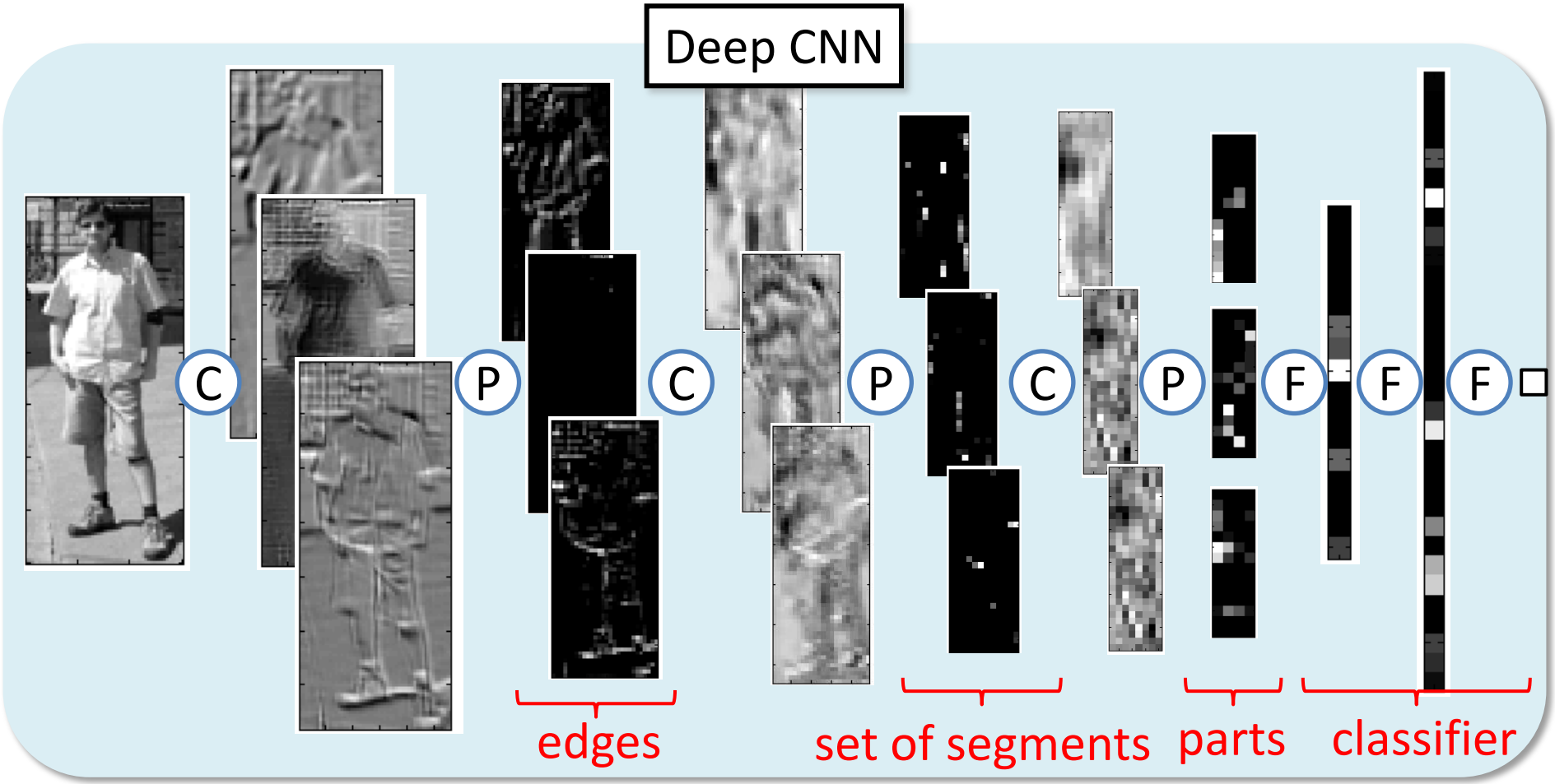
- **Convolutional Neural Network** (CNN) comprising
  - **Convolutional** layer(s) → local feature extraction
  - **Pooling** layer(s) → dimensionality reduction
  - **Fully-connected** layer(s) → classification/regression

visual features automatically extracted



Why is “Deep” Architecture Advantageous?

- Transforming data vectors at multiple layers help to increase the level of abstraction of visual contents (Bengio, 2009).



- Why Deep Neural Networks?
- **Why GPGPUs?**
- Demo Setup

Major Reason

- **Classification** can be boosted.
  - Deep CNN can run at real-time with a compact Tegra K1 processor.

why? ➡ *A Deep CNN typically has a **small** number of sequential operations, each of which is dominated by **highly-parallelizable**, massive sums-of-products.*

Deep CNN is very GPU friendly!

Minor Reasons

- **Classification** can be boosted.
  - Deep CNN can run at real-time with compact Tegra K1 processor.
- **Learning** can be boosted.
  - It can take a month without GPUs. Takes only several days with GPUs.
- **Development** can be boosted.
  - Capability of processing generic programing language (i.e., CUDA) makes R&D process efficient.
  - No more fixed-points!

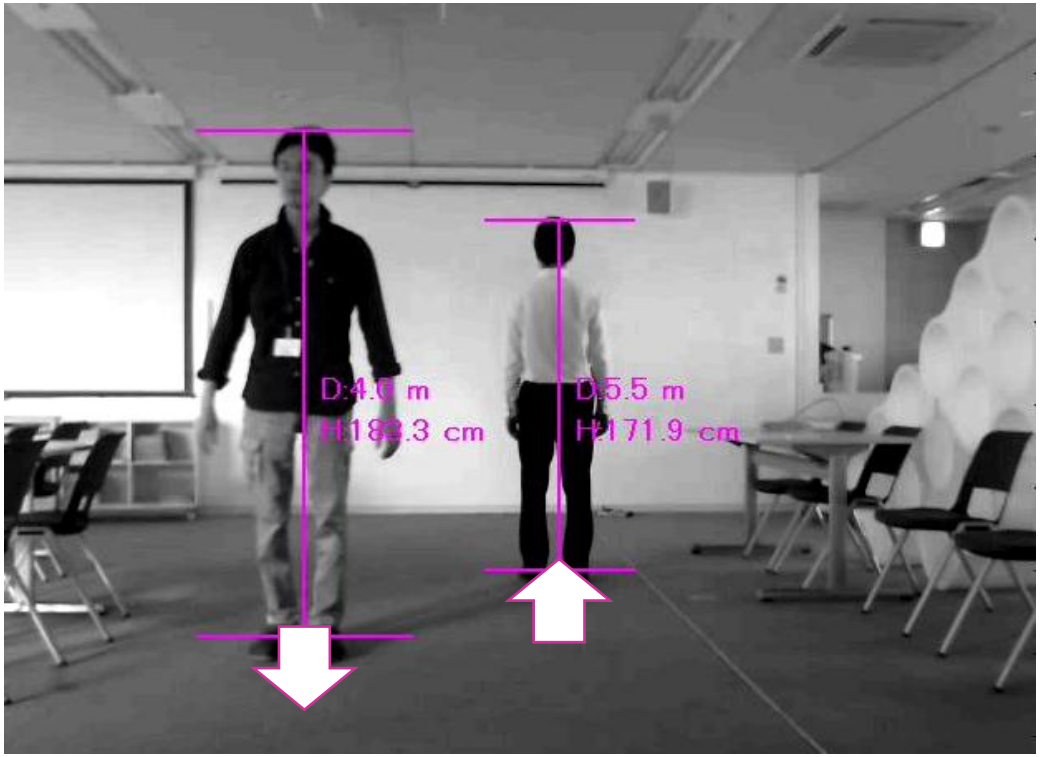
Minor Reasons

- **Classification** can be boosted.
  - Deep CNN can run at real-time with compact Tegra K1 processor.
- **Learning** can be boosted.
  - It can take a month without GPUs. Takes only several days with GPUs.
- **Development** can be boosted.
  - Capability of processing generic programing language (i.e., CUDA) makes R&D process efficient.
  - No more fixed-points!

- Why Deep Neural Networks?
- Why GPGPUs?
- **Demo Setup**

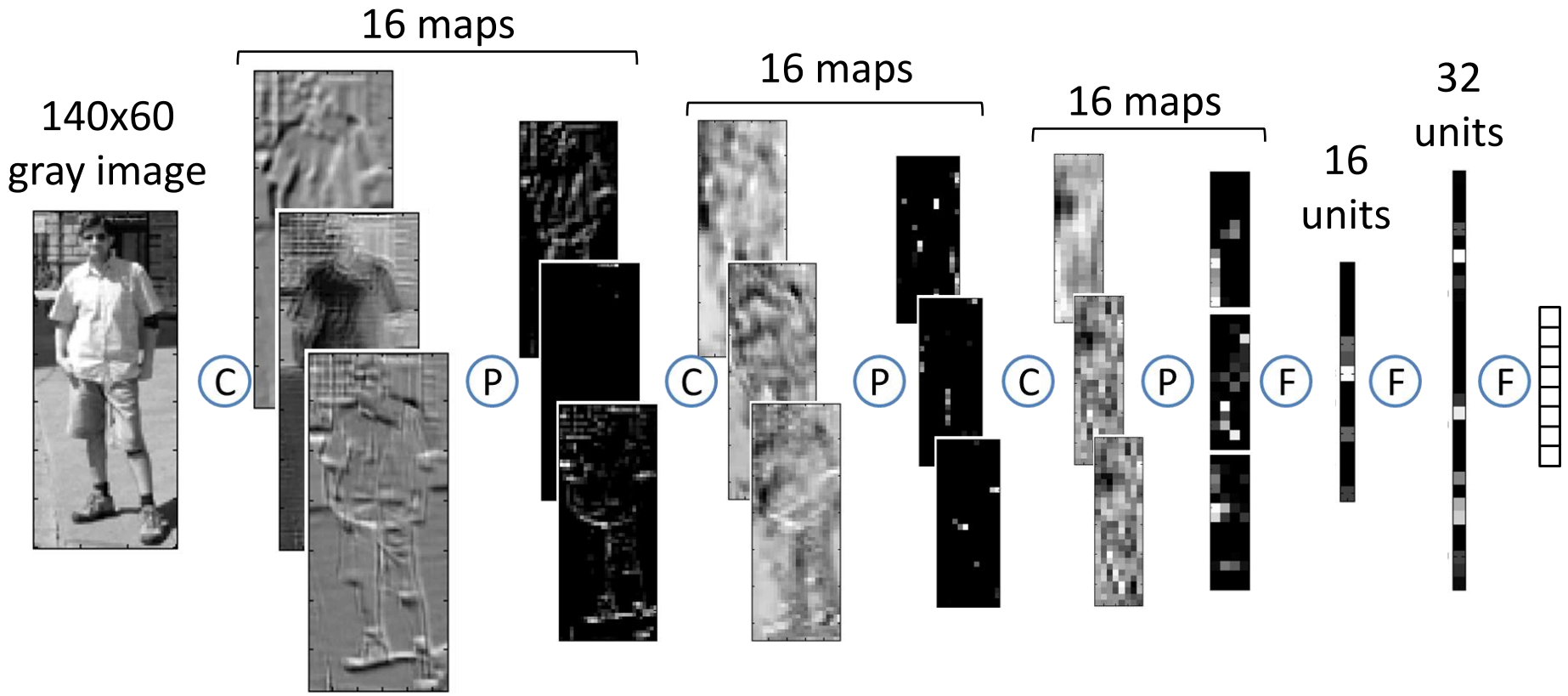
Problem Statement

- Real-time pedestrian **detection** with **depth**, **height**, and **body orientation** estimations
  - requires ***discrete***-valued classification and ***real***-valued regression tasks



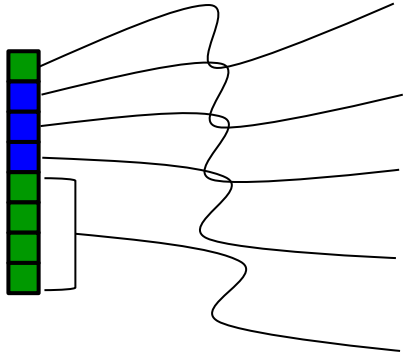
Deep CNN Model

- A single deep CNN with 9 (processing) layers
  - 3 Convolutional + (Max) Pooling layers
  - 3 Fully-connected layers



Deep CNN Model

- Output layer contains:



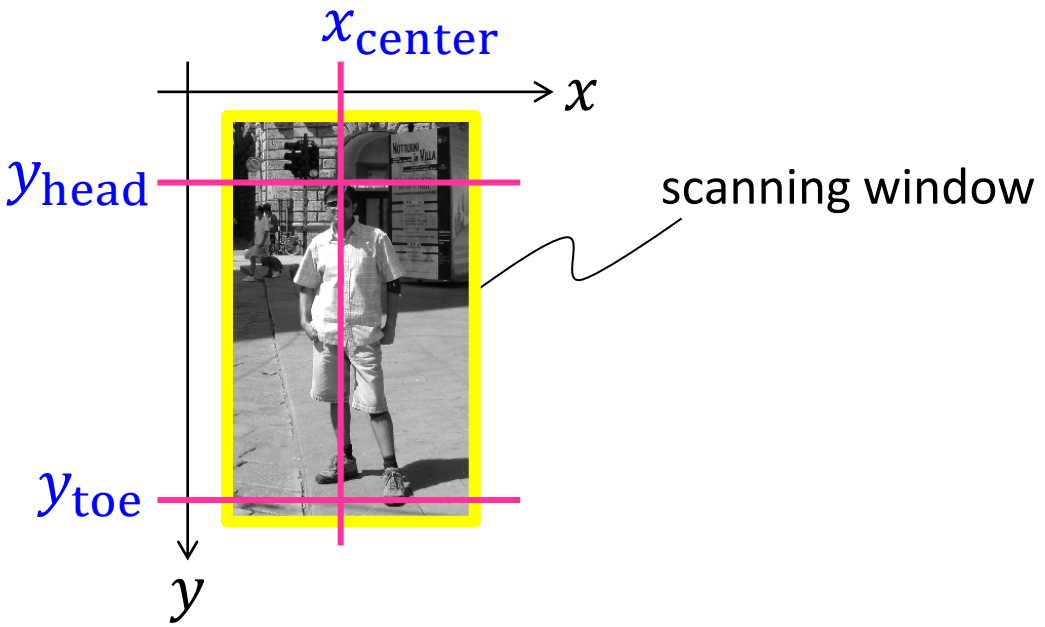
1 binary classification unit for pedestrian/background

1 regression unit for relative x-coord. of the body center

1 regression unit for relative y-coord. of the top of the head

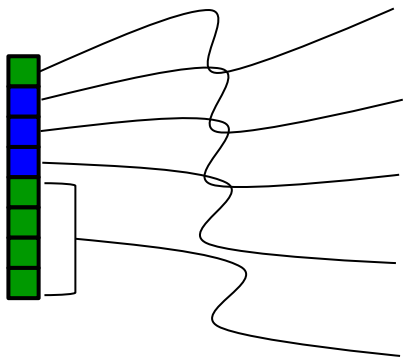
1 regression unit for relative y-coord. of the toe

4 classification units for body orientations



Deep CNN Model

- Output layer contains:



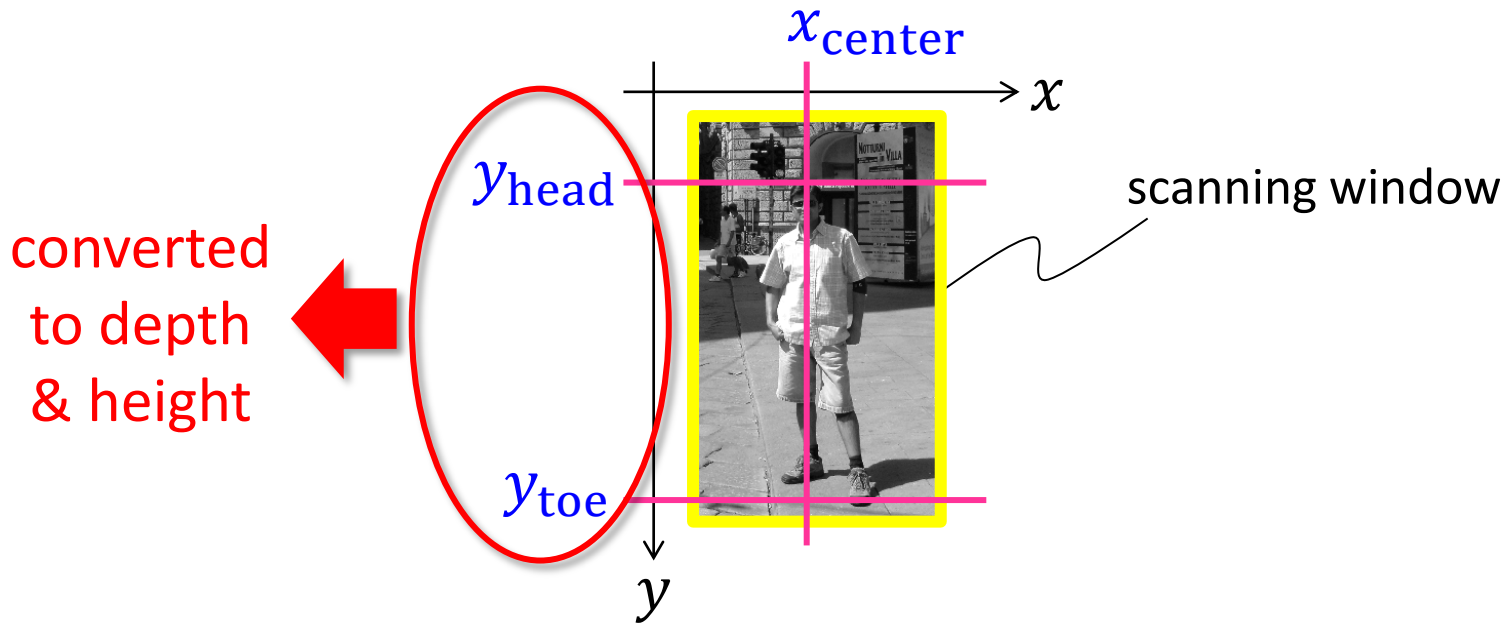
1 binary classification unit for pedestrian/background

1 regression unit for relative x-coord. of the body center

1 regression unit for relative y-coord. of the top of the head

1 regression unit for relative y-coord. of the toe

4 classification units for body orientations



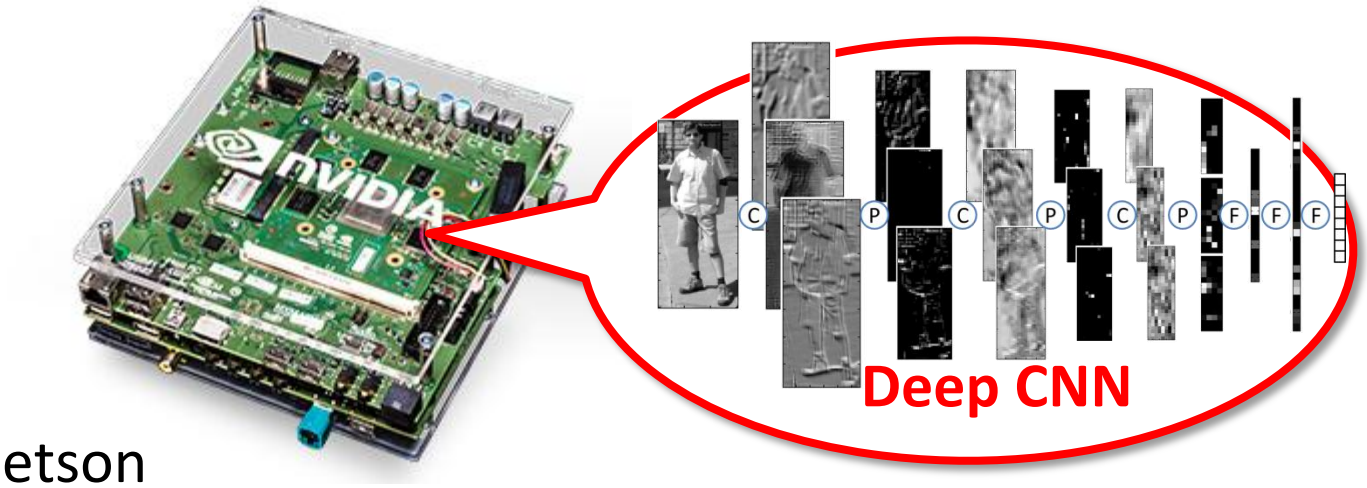
Processor

- NVIDIA's Jetson platform equipped with **Tegra K1** processor
  - 192 Kepler CUDA cores
  - 200 GFLOPS (automotive K1)
  - memory shared between GPU and CPU (ARM)
  - low power consumption for automotive applications, e.g., ADAS

Jetson

Processor

- NVIDIA's Jetson platform equipped with **Tegra K1** processor
  - 192 Kepler CUDA cores
  - 200 GFLOPS (automotive K1)
  - memory shared between GPU and CPU (ARM)
  - low power consumption for automotive applications, e.g., ADAS



Jetson

Take-Home Messages

1. Extracting rich information about pedestrians will be one of the key technologies for automotive safety.
2. Deep Neural Networks are the best algorithms for visual recognition tasks.
3. GPGPUs enable real-time processing.
   - *Demonstrated that Tegra K1 indeed does!*