

Deep Learning Processing Technologies for Embedded Systems October 2018

Neural Networks Architecture





Multi-Task Vehicle Detection With Region-of-Interest Voting

Multi Task NN



Popular Computer Vision Tasks

Object Classification



Object Detection



Semantic Segmentation



Case Study: Rear-View Camera

Use Case I: Rear-Collision Warning (RCW)





Use Case II: Lane-Change Assist (LCA)

Use Case III: Parking Assist (PA)





Use case Example: Rear-Collision Warning

5	
5	

Scenario

	and the second	
Environment:	High-speed cruisir	ıg
Hovering speed (Rear car):	120 Km/H	
Max Closing speed:	80 Km/H	
Typical acceleration:	3 m/s ²	
Viewing angle:	120 deg	
	and the second se	



System Requirements	
Policy:	Always-on
Strategy:	Match speed using ACC
Safety margin:	2 sec





Use case Example: Rear-Collision Warning





Use case Example: Rear-Collision Warning

75.1%



Neural Network Model Definition

•	Task:	Object Detection
•	Model:	SSD detector
•	Feature extractor:	VGG-16
	Compute per frame:	~106 GMACs

- mAP (VOC2007):
- Frames for 1 detection (5σ)

Compute Requirements @ 20 fps

2.1 TMAC

	-
E	~
Ę.	<u> </u>
	_



https://arxiv.org/abs/1512.02325



Theoretical Efficiency







Real-world Processors Efficiency for Deep Learning Image Processing



Neural Networks - Observations

Control

- Flexibility during compile time
- Fully deterministic in runtime

Memory

- Parameters and partial sums are localized
- Layer outputs move around (but not too far)

Compute

- Recurring operations (MACs >> Activations)
- Mostly low precision

Interconnect







Neural Networks - Resource Balance



Memory, Control and Compute balance changes dramatically along the network's layers



Approaches To Deep Learning Processing

Fixed Function Accelerator

- Theoretically optimal at a specific workload
- Minimal flexibility

Von-Neumann Architecture

- Temporal resource allocation
- Common memory space
- Classical flow control programming model

Symmetric Dataflow Architecture

- Spatial resource allocation
- Segregated memory spaces
- Balanced graph oriented









Von-Neumann Architecture





Symmetric Dataflow Architecture





Real-world Processors Efficiency for Deep Learning Image Processing



Use case Example Summary: Rear-Collision Warning

50 m Detection Distance



O.3MP Camera Resolution

20 fps Frame Rate





Thank You!

contact@hailotech.com

Icons from icons8.com