

Deep Learning Inference using FPGA

- FPGA를 활용한 딥러닝 추론 구현:
LeNet and Yolo -

2020

Ando Ki, Ph.D.

adki@future-ds.com

Goals and objectives

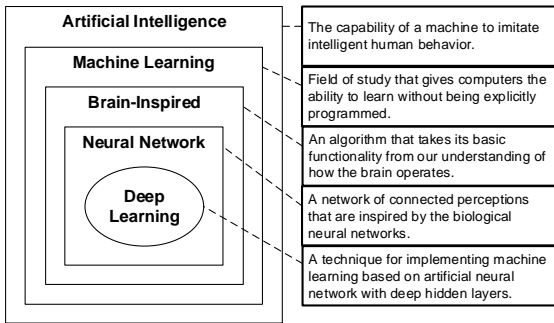
■ Goals

- ▶ Understanding of artificial intelligence, machine learning, and deep learning.
- ▶ Acquiring the working knowledge of deep learning model.
- ▶ Practicing development and running deep learning model.

■ Objectives

- ▶ Understanding of deep neural network
- ▶ Understanding of well known DNN for image classification.
- ▶ Understanding of deep learning frameworks: TensorFlow, Caffe V1, PyTorch and so on
- ▶ Understanding of LeNet
- ▶ Understanding of Darknet/YOLO
- ▶ Understanding of light deep learning networks: SqueezeNet, ZynqNet, MobileNet, and so on (not included yet)

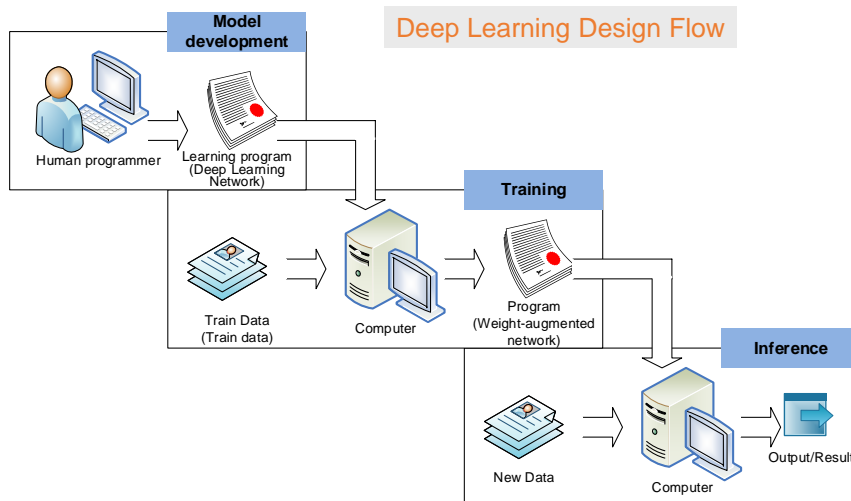
What have been covered



- AI (인공지능)
 - ▶ 지능적인 사람의 행동/판단을 모사하는 기계의 능력
- ML (기계학습)
 - ▶ 명시적인 프로그래밍하지 않고 컴퓨터가 배우는 능력에 대해 공부하는 영역
- DL (딥러닝)
 - ▶ 많은 수의 숨은 레이어를 갖는 인공 뉴런 네트워크에 기초한 기계학습의 구현

(3)

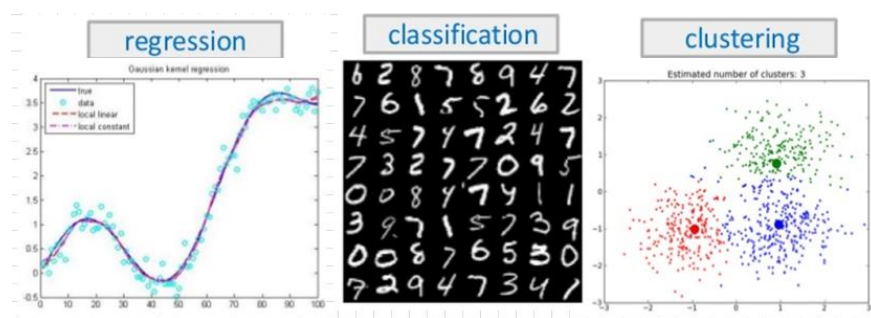
What have been covered



(4)

What have been covered

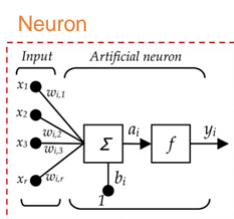
Regression, classification, clustering



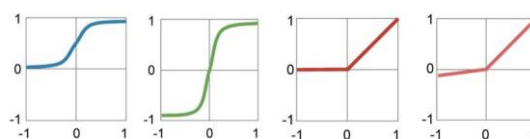
(5)

What have been covered

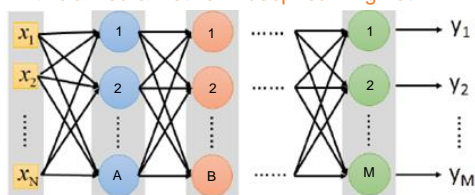
Deep-learning related topics



Activation functions



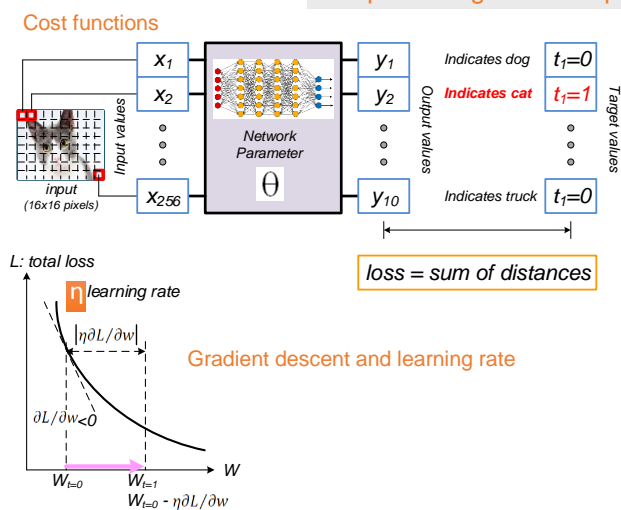
Artificial neural network: deep-learning net



(6)

What have been covered

Deep-learning related topics



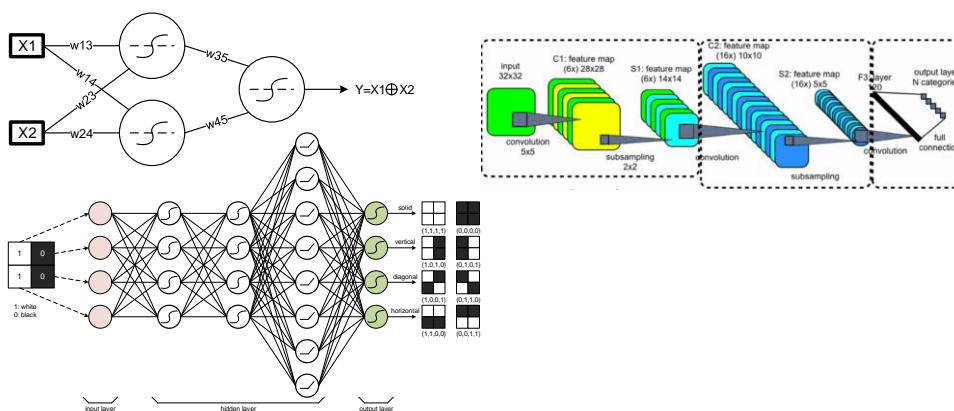
Softmax
CNN
Covolution
Pooling

(7)

What have been covered

Tiny-Dnn

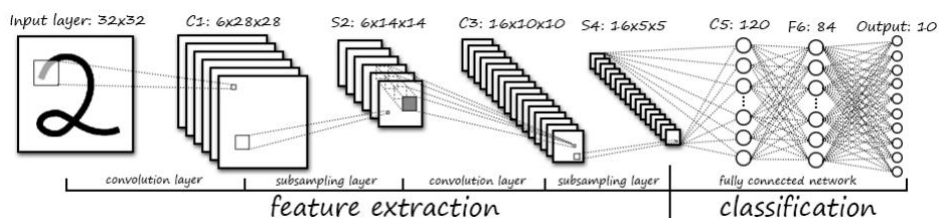
- header only, dependency free deep learning library written in C++ with C++14 features -



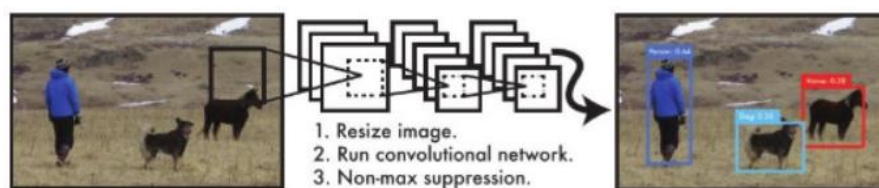
(8)

What have been covered

LeNet

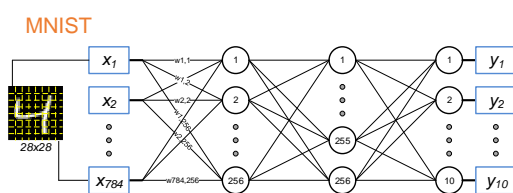


YOLO



9

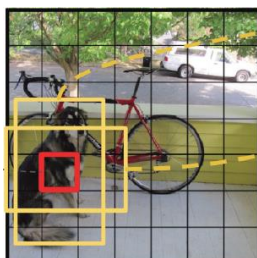
What have been covered (optional)



TensorFlow

TensorFlow

YOLO



Caffe

(10)

Running examples

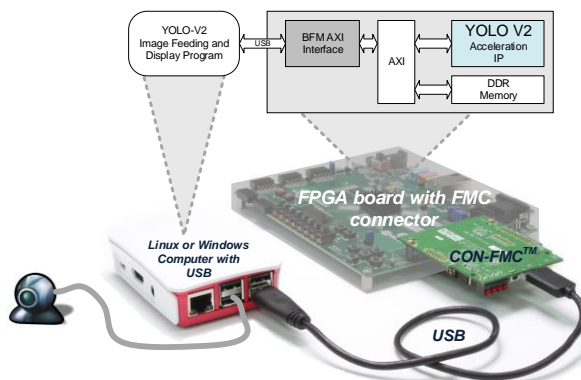
■ High-performance

- ▶ PCI-Express FPGA card
- ▶ Refer to <http://www.future-ds.com/en/products.html#DeepAccel-DualVU9P>



■ Low/medium-performance

- ▶ USB connection card
- ▶ Refer to video clips on <http://www.future-ds.com>



www.FUTURE-DS.com

11

Benefits/Advantages of FPGA on Deep Learning

- FPGAs have product lifecycles of 15 years.
- High performance per Watt and low latency make it suitable for real-time embedded applications.
- The FPGA logic can be shaped to match any network architecture.
- Performance, cost and power will define the FPGA of choice.
- Future proof and scalable solution as the FPGA architecture can be re-configured for future neural networks.
- The deep learning core can be easily integrated with other CPU's, vision functionality and connectivity.
- Future Design Systems framework offers a flexible approach to program the FPGA and a fast-time to market.

12

(주)퓨처디자인시스템

34051 대전광역시 유성구 문지로 193, KAIST 문지캠퍼스, F723호
(042) 864-0211~0212 / contact@future-ds.com / www.future-ds.com

Future Design Systems, Inc.

Faculty Wing F723, KAIST Munji Campus, 193 Munji-ro, Yuseong-gu, Daejeon 34051, Korea
+82-042-864-0211~0212 / contact@future-ds.com / www.future-ds.com



FUTURE
Design Systems