Day 1: Introduction to AI

Part 2: Quick Tour of Deep Learning

2021

Part 2:
Quick Tour of Deep Learning

Content

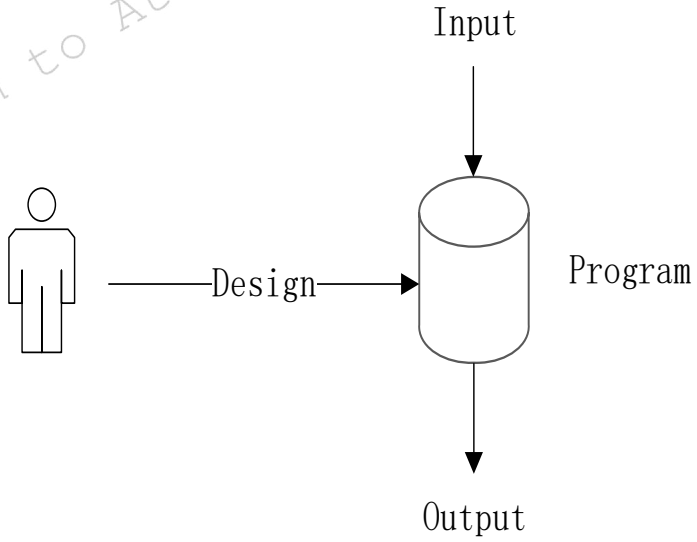
1.2 Quick Tour of Deep Learning

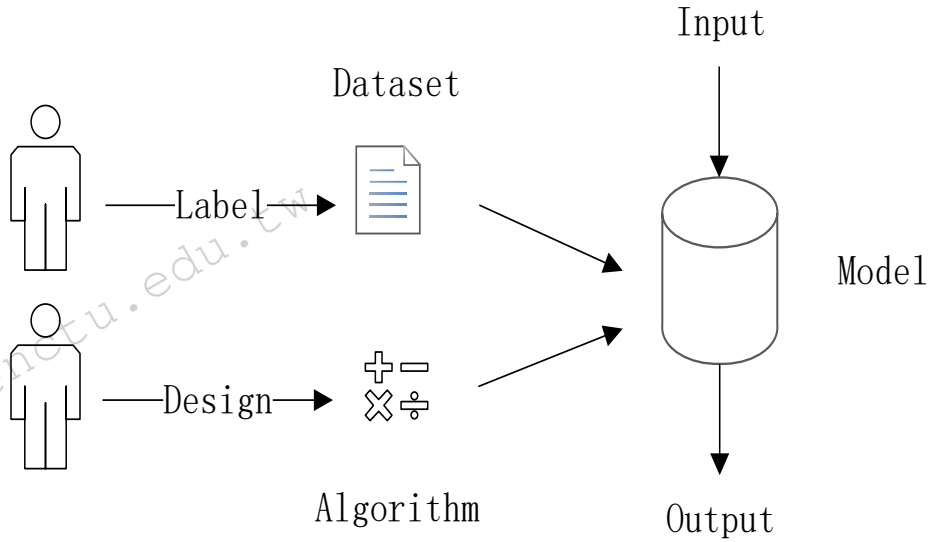
◆ 1.2.1 From ML to DL: What is Deep learning?

- Traditional programming vs. ML
- Summary: Paradigm of ML
- Why Deep Learning?

Traditional programming vs. ML

- Traditional programming defines patterns (*program*) by hands, and it is used to process input and to generate output.
- In training stage, the patterns (*model*) in ML is learned from labeled data (including examples and desired labels, or input and wanted output), and ML use this model to process new data during inference stage.

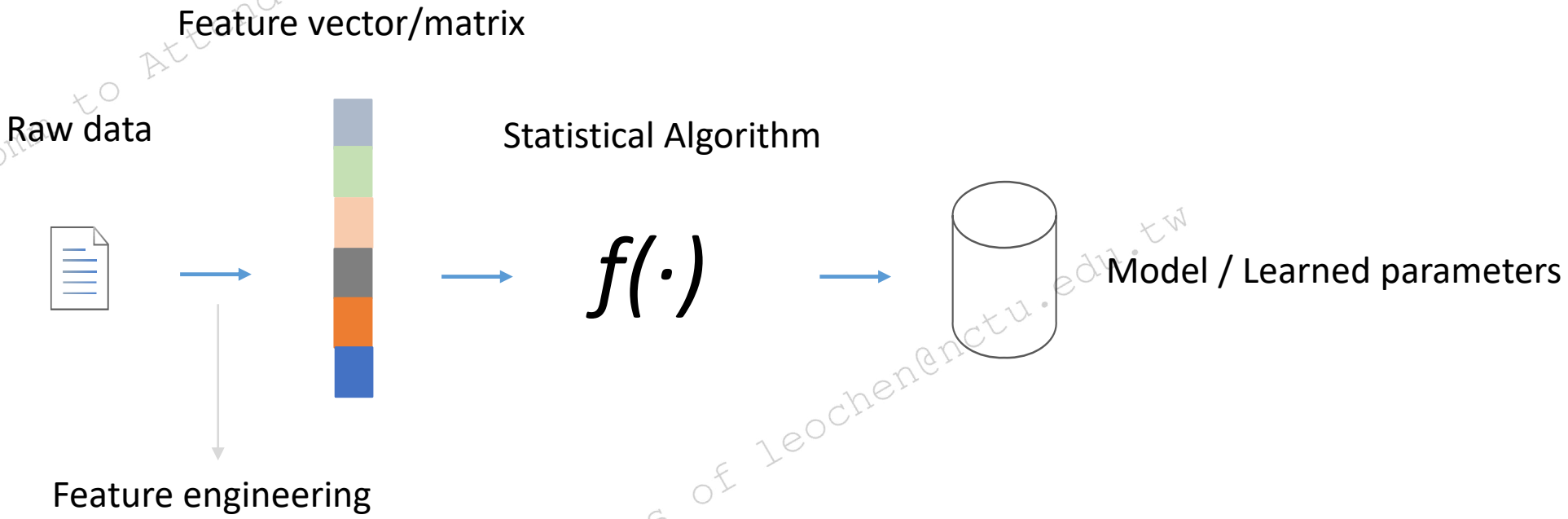




Qualcomm

Summary: Paradigm of ML

- Paradigm of ML = Feature engineering + Statistical Algorithm
- Methods for Feature engineering are complicated well-designed mathematical formula, while the application scopes of these formulas are not universal.



Qualcomm

Examples of feature engineering

- Feature engineering is the process of using **domain knowledge** of the data to create features that make machine learning algorithms work.
- Typically, different task in machine learning has different feature engineering method:

Haar-like features[1] for face detection

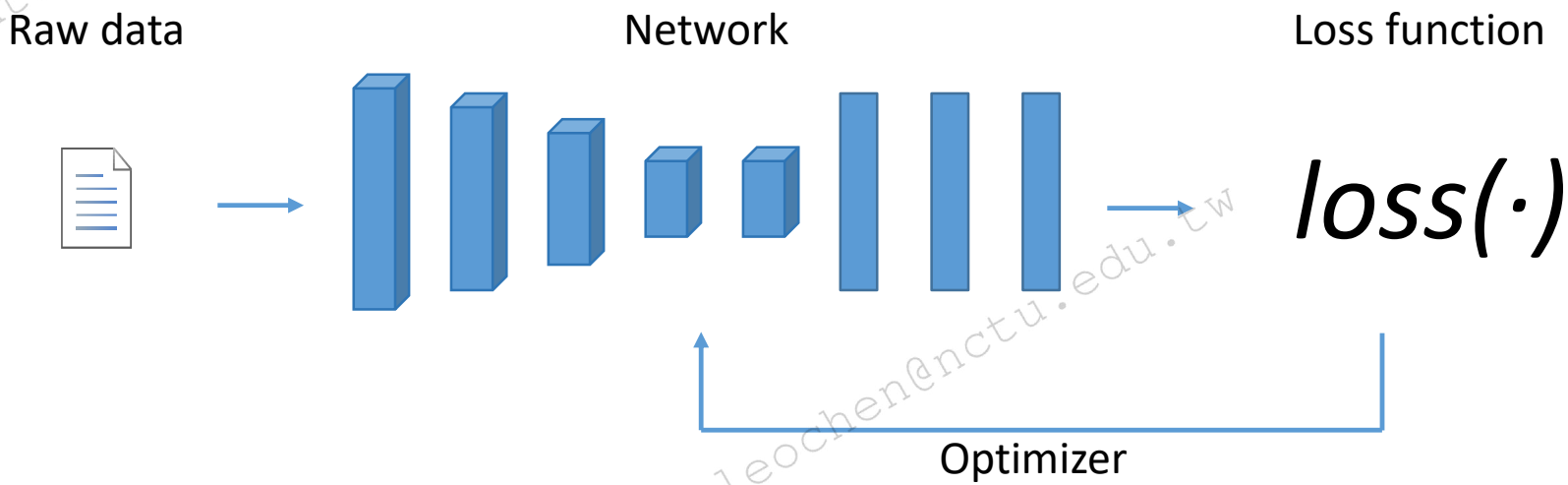
HoG feature[2] for human detection

Optical flow[3] for motion detection

[1] P Viola, M Jones, Rapid Object Detection using a Boosted Cascade of Simple Features, *CVPR'01.*
[2] N Dalal, B Triggs, Histograms of Oriented Gradients for Human Detection, *CVPR'05.*
[3] Horn B K P, Schunck B G. Determining optical flow[J]. Artificial intelligence, 1981,

Why Deep Learning? It removes feature engineering!

- Paradigm of DL = Design network structure and other hyper-parameters + Raw data



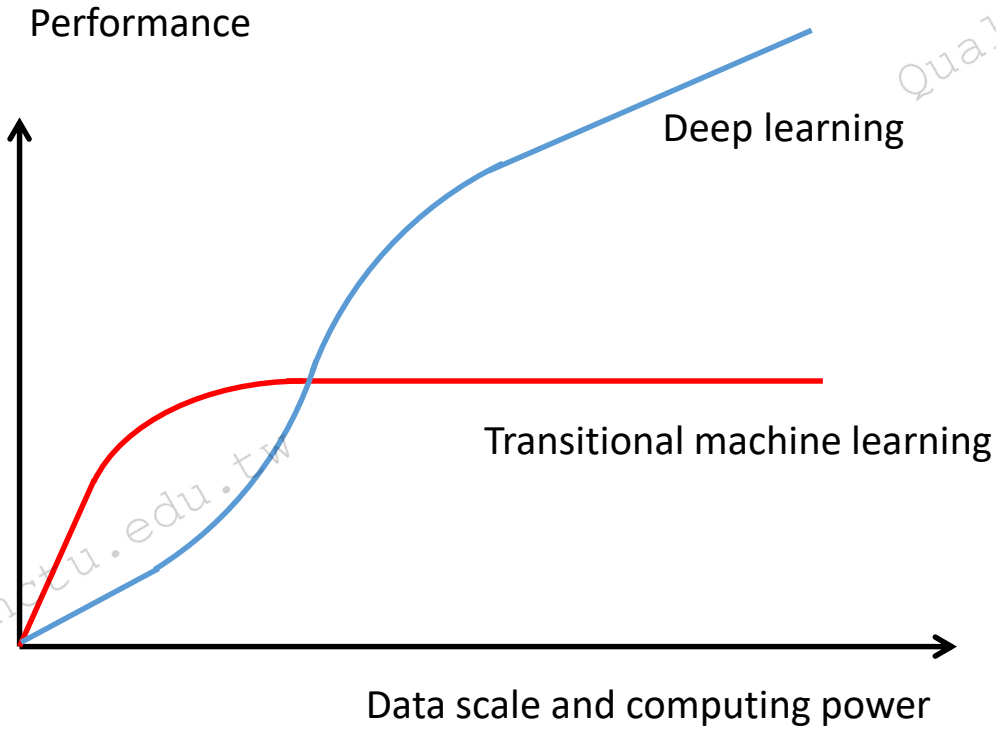
Qualcomm

Why Deep Learning? It performs better than other models!

- Deep learning = **A**lgorithm + **B**ig data + **C**omputing

In the early years, computing power and data volume were insufficient. Artificial neural networks (ANN) algorithms comparable or worse than traditional algorithms.

Why did deep learning suddenly break out in recent years? This is due to the explosion of data in the era of big data as well as the explosion of computing power benefit from by Moore's Law. The performance of deep learning exceeds traditional methods by far.



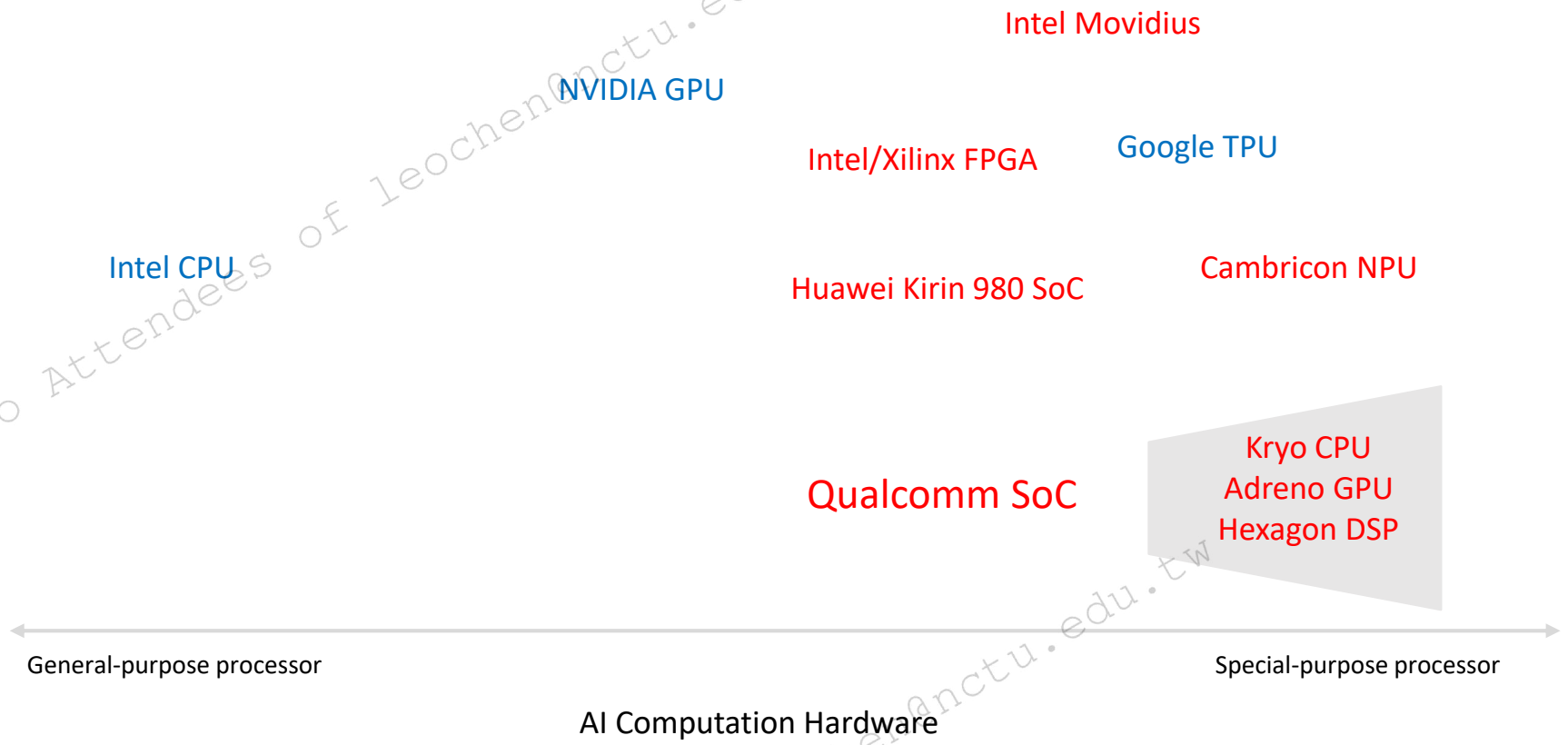
Qualcomm

1.2 Quick Tour of Deep Learning

◆ 1.2.2 Deep learning infrastructure: from hardware to software

- Chipsets for Training and Inference: CPU, GPU, FPGA, ASICs
- Running Platform: PC, Cloud, Device
- Framework of Deep Learning
- Inference engine of Deep Learning

Chipsets for Training and Inference: CPU, GPU, FPGA, ASICs



Accelerator for inference only
Accelerator for both training and inference

Qualcomm

Chipsets for Training and Inference: CPU, GPU, FPGA, ASICs

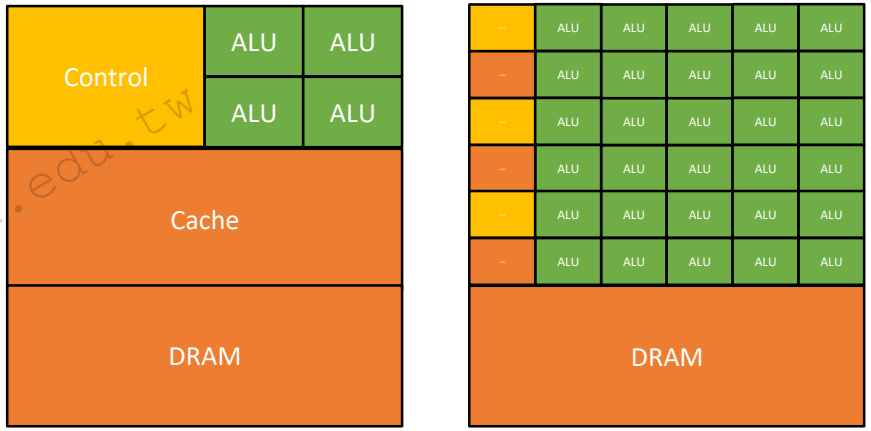
Generally speaking, processors can be divided into:

- **General-purpose processor,** generally called Central Processing Unit (CPU), performs all general tasks in computer.
- **Special-purpose processors** (Application-specific integrated circuit, ASICs, or Coprocessor) are designed for a specific application.
- **Field Programmable Gate Arrays (FPGAs)** are a special type of ASICs. Based on a matrix of configurable logic blocks (CLBs) connected via programmable interconnects, FPGAs can be reprogrammed to desired application or functionality requirements after manufacturing.

Chipsets for Training and Inference: CPU, GPU, FPGA, ASICs

CPU, GPU, FPGA, ASICs with deep learning

- CPU: CPU can run basic training process indeed, but it is really slow.
- GPU: GPU has more ALU than CPU, so its parallel computing power is hundreds of times faster than CPU for compute-intensive task like training.
- FPGAs have been recently adopted for accelerating the implementation of deep learning networks due to their ability to maximize parallelism as well as due to their energy efficiency.
- Application-specific integrated circuit (ASICs) for AI: An integrated circuit (IC) customized for a particular use (model training and inference here), including TPU/NPU/Cloud AI 100.



Qualcomm

Running Platform: PC, Cloud, Device

	PC	Cloud	Device
Training	The most common computing device.	1. Strong computation power. 2. Good scalability. 3. Easy access to big data.	
Deploy	The most common computing device.	1. Strong computation power. 2. Good scalability. 3. Suitable for API services. *Deploy once, called anywhere.*	1. Friendly to privacy, security 2. Quick response without network connection. 3. Small and easy to carry.

Frameworks of Deep Learning

- Frameworks of deep learning help developers quickly train and deploy a deep learning model, they provide a set of commonly used function libraries and class libraries for deep learning, and builds a deep network model just like "building blocks."

Framework	Author	Year	Open source	license	Written in	Interface
Caffe	Berkeley	2013	Yes	BSD	C++, CUDA	Python, MATLAB, C++
TensorFlow	Google	2015	Yes	Apache 2.0	C++, Python, CUDA	Python, C/C++, Java, Go, etc.
MXNet	Apache Software Foundation	2015	Yes	Apache 2.0	Python, C, C++, CUDA	C++, Python, Java, Go, R, etc.
PyTorch	Facebook	2016	Yes	BSD	Python, C, C++, CUDA	Python, C++
Torch	Ronan Collobert, et al.	2002	Yes	BSD	C, Lua	Lua, C
Theano	University of Montreal	2007	Yes	BSD	Python	Python
MATLAB	MathWorks	-	No	-	C, C++, Java, MATLAB	MATLAB
Keras	François Chollet	2015	Yes	MIT license	Python	Python, R
CNTK	Microsoft Research	2016	Yes	MIT license	C++	Python, C++

Qualcomm

Inference engines of Deep Learning

- Unlike the deep learning frameworks, which focus mainly on training and deploying on CPU and GPU devices, the inference engine emerges to help models to run on heterogeneous computing platforms (including servers and PCs, mobile devices, DSPs). Tools such as performance optimization, model encryption, and memory analysis are provided.

Inference engine	Author	Year	CPU	GPU	DSP	NPU / VPU
SNPE	Qualcomm	2016	1	1	1	1
TensorRT	NVIDIA	2016	0	1	0	0
OpenVINO	Intel	2018	1	1	0	1
MACE	Xiaomi	2018	1	1	0.5	0
TensorFlow Lite	Google	2017	1	0.5	0	0
Core ML	Apple	2018	1	1	0	1
ONNX	Microsoft/Facebook/Amazon	2017	1	1	0	0

Qualcomm

1.2 Quick Tour of Deep Learning

◆ 1.2.3 "Textbook" of Deep learning: Datasets

- Public dataset: ImageNet, Pascal VOC, COCO and so on
- Custom dataset: Data collection and labelling method

Public dataset: ImageNet, Pascal VOC, COCO and so on

- Image Classification: ImageNet
- Object Detection: Pascal VOC, COCO
- Face Detection: FDDB

- Notice: Licenses!

Custom dataset: Data collection and labelling method

- If public datasets can't satisfy your needs, or the license is not suitable for your application scenario, you need to collect your own dataset. This is some important considerations:

1. **Privacy and legal.** Please to abide by the law and the license, pay attention to privacy issues.
2. **Cost Control. A**lgorithms are open published, **C**ommuting equipment is a one-time investment, but **B**ig data is indeed expensive.
3. **Quality assurance.** If the input data has wrong or inaccurate labels, the trained model definitely will not work well, even performance metrics are also untrusted. "Garbage in, garbage out." Imagine a lot of mistakes in textbooks and exam answers?
4. **Don't reinventing the wheel.** Using open source annotation tools, instead of redevelopment, this may save time and money.

Custom dataset: Data collection and labelling method

- Let's say, you want to develop an application to detect and predict corresponding categories of fruits by camera. You need to train an object detection model for different fruits, like watermelon, apple, grape, etc.
- But you can't find a public dataset about fruits or containing well-labeled fruits' location. You think about creating a dataset by yourself.

1. **Privacy and legal.** You can take photos of own fruits in supermarket with the consent of the staff. Or you can find some image about fruits from other public dataset or the internet if your purpose (commercial or research) meets the copyright requirements.
2. **Cost Control.** Supposed that you collect a large number of unlabeled images (e.g. 20k images). You so busy to develop program that you have to assign annotation tasks to others and pay them money.

Custom dataset: Data collection and labelling method

3. **Quality assurance.** Humans always make mistakes. In order to make sure the annotation data is exactly correct, you can ask two people to label the same data separately, compare their output, and re-label if two annotations of the same image are too different. (Of course, double the cost.)

4. **Don't Reinventing the wheel.** Use labelImg to label objects in images to train your own detection model.

https://github.com/tzutalin/labelImg

Qualcomm

1.2 Quick Tour of Deep Learning

◆ 1.2.4 History, Present and Future

- History of Deep Learning
- Heroes of Deep Learning
- Present of Deep learning
- Limitations of Deep Learning
- Future of Deep Learning

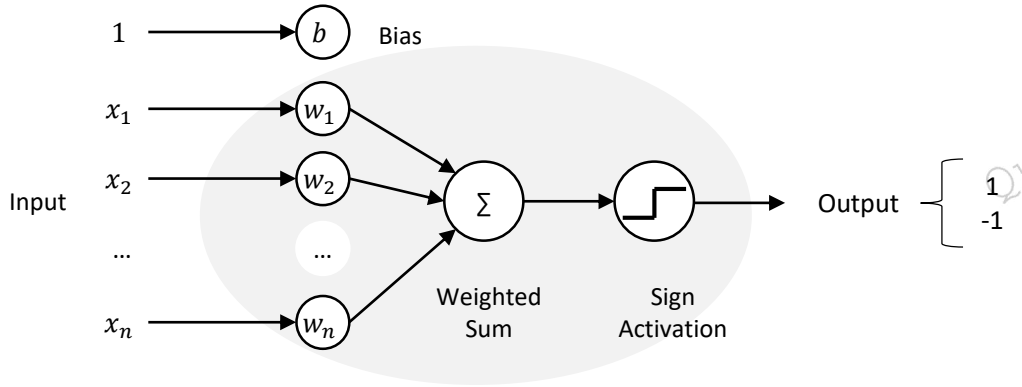
History of Deep Learning

- The term ***Deep Learning*** was introduced to the machine learning community by Rina Dechter in 1986.
- Deep learning, relative to “Shallow learning”, represents the number of layers through which the data is transformed. A “deep” model is consisted with several “shallow” models.
- In “ancient times” of deep learning, it also includes other layer-wise models like deep belief networks (DBN) and deep Boltzmann machines (DBM); but modern Deep Learning usually refers to the improved version of artificial neural networks (ANN).
- Throughout the history of artificial intelligence, we are at the third wave of deep learning (or neural networks).

History of Deep Learning

The first wave:

- The prototype of deep learning appears in ***cybernetics***. Rosenblatt[1] (1958) created the perceptron, an algorithm for pattern recognition. It's seen as the simplest neural network.
- Neural network research stagnated after by Minsky and Papert (1969)[12], they pointed out two key issues of neural network at the time:
- 1. Perceptron was incapable of processing the exclusive-or circuit(XOR).
- 2. Computers didn't have enough processing power to effectively handle the work required by large neural networks.



Artificial intelligence researchers focused on high-level (symbolic) models like expert systems with knowledge embodied in ***if-then*** rules in that time.

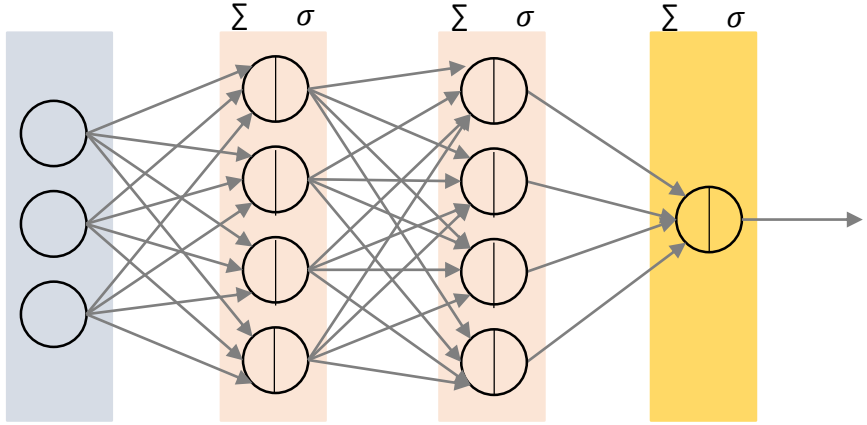
Example of Perceptron on XOR

[1] Rosenblatt, F. (1958). "The Perceptron: A Probabilistic Model For Information Storage And Organization In The Brain". Psychological Review. 65 (6): 386–408.
[2] Minsky, Marvin; Papert, Seymour (1969). Perceptrons: An Introduction to Computational Geometry

History of Deep Learning

The second wave:

- Werbos (1975) proposed ***backpropagation*** algorithm that made the training of multi-layer networks feasible and efficient. The term ***connectionism*** became popular to describe artificial neural networks (ANN) in the mid-1980s.
- However, kernel methods like Support vector machines (SVM) and other much simpler methods such as linear classifiers exceeded finally in 1990s, because ANN didn't have much advantage than other models.
- In the next period of time, some researchers continued to study in this field. In 1992, max-pooling in CNN was introduced to help with computer vision; Schmidhuber adopted recurrent neural network (RNN) as long short-term memory (LSTM) in 1997; LeNet-5 is proposed in 1998.



Questions:
Combine the characteristics of deep learning mentioned above, why the first two waves finally receded?

Example of MLP on XOR

Qualcomm

History of Deep Learning

The third wave:

- Hinton et al. (2006)[1] proposed a many-layered feedforward neural network **Deep Belief Network**, and a training strategy named greedy layer-wise pre-training.
- In 2012, Google X, led by Andrew Y. Ng and Jeff Dean, detected objects (e.g. cats and human faces) from YouTube videos with neural network model trained by 16000 computers.
- In the same year, AlexNet won the ILSVRC 2012 competition by a larger margin (error rates: 15.3% of AlexNet VS. 26.2% of 2nd place). ResNet by MSRA got a top-5 error rate of 4.94% and surpassed humans (estimated 5.1%) in 2015.
- In 2017, AlphaGo, the successor of AlphaGo Master, beat Ke Jie, the world No.1 ranked player at the time.

[1] Hinton, G. E.; Osindero, S.; Teh, Y. (2006). "A fast learning algorithm for deep belief nets" (PDF). Neural Computation. 18 (7): 1527–1554.
[2] Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification, arxiv 1502

Heroes of Deep Learning

- Geoffrey Hinton: BP algorithm, Boltzmann machine, contributions to CNNs
- Yann LeCun: CNN, improving BP
- Yoshua Bengio: RNNs, neural language model, GANs
- Jürgen Schmidhuber: Sophisticated versions of RNN called LSTM
- Jeff Dean: Leader of Google AI, TensorFlow
- Andrew Ng: Leader of Google Brain, Baidu IDL, Founder of Coursera
- Fei-Fei Li: ImageNet, Google Cloud AI
- Ian Goodfellow: GANs (student of Bengio)
- …

} Turing Award 2019

Present of Deep learning

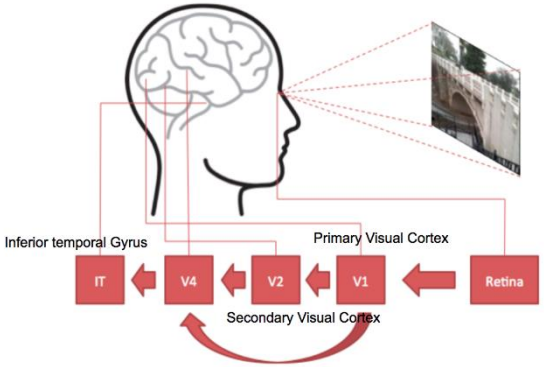
- We will quick view three types of famous modern deep learning models.
  - Convolutional Neural Network (CNN)
  - Recurrent Neural Network (RNN) & Long short-term memory (LSTM)
  - Generative adversarial network (GAN)

Qualcomm

CNN

We will introduce it in detail in next lesson.

- The design ideas were inspired by visual processes in biological brains. Convolution, non-linear activation function and pooling are basic blocks of CNNs.
- LeNet-5, a pioneering 7-level convolutional network by LeCun et al. in 1998, was used to recognize hand-written numbers on checks in 32*32 pixels images. Modern deep learning models are significantly affected by design ideas in LeNet-5.
- CNN is widely used in computer vision fields, like image classification, image regression, object detection, image segmentation and so on.

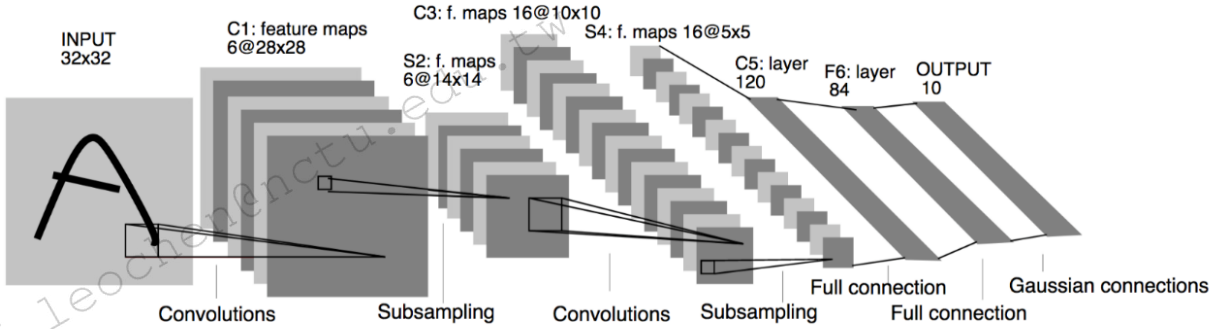


V1: Edge detection, etc.

V2: Extract simple visual properties (orientation, spatial frequency, color, etc)

V4: Detect object features of intermediate complexity

TI: Object recognition.

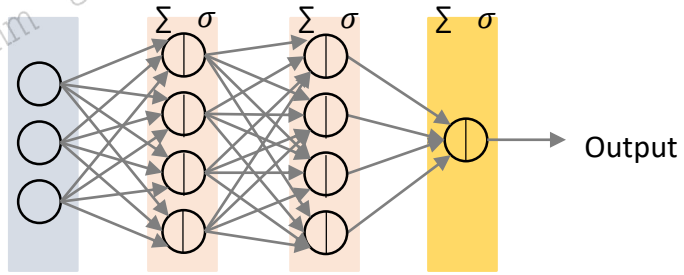


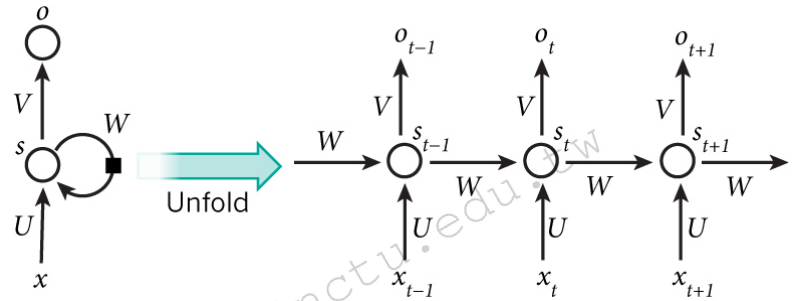
On the Origin of Deep Learning. arxiv 1702.07800
Gradient-based learning applied to document recognition, 1998

RNN

Re-think:

- In traditional multi-layer perceptron (MLP), a neuron in a layer is fully connected with another neurons in the next layer. But neurons from the same layer are connectionless with each other.
- In natural language processing (NLP) tasks, it is necessary to process **time series** data. For example, word spelling correction requires knowledge of the context before and after the word.
- Recurrent Neural Networks was so named because the hidden unit of RNN has **a connection to itself**, so it can be expanded to represent the input at a certain moment and the intermediate state at the previous moment determines the current output.





This structure gives the RNN "memory" function to “remember” the context in NLP tasks.

LeCun, Bengio, and Hinton, 2015; Deep learning

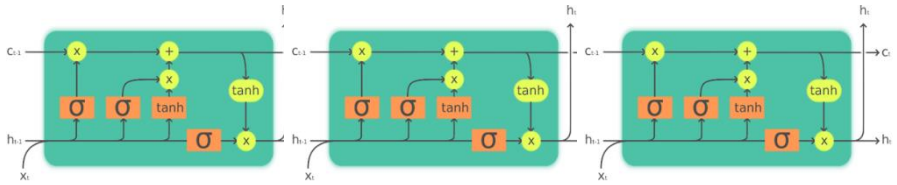
Please giv me a hand. → Please give me a hand.

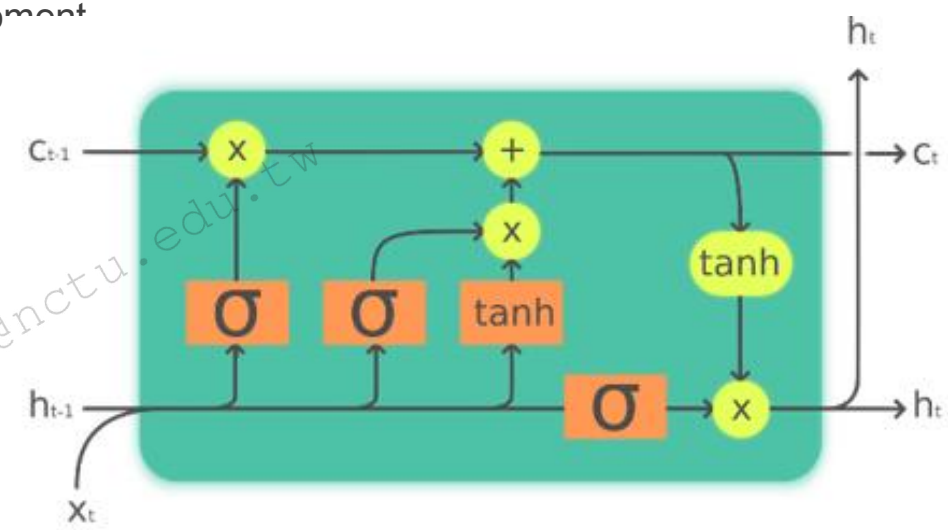
Word spelling correction

Qualcomm

RNN / LSTM

- Studies have shown that RNN does not remember long-term information very well. Long-Short Term Memory networks (LSTM) is a variant of RNN that remembers information for long periods of time.
- There are two states in LSTM: cell state and hidden state. Here are main phases within LSTM:
  1. Forget stage. Selectively forgetting the input x_i and output of previous node h_{i-1}. *"Forget the unimportant."*
  2. Generate new memory stage. A sigmoid and a tanh is used to generate a value to remember. *"Remember the important"*
  3. Update memory stage. Update the cell stage based on the “left memory” (stage 1) and the “new memory” (stage 2).
  4. Output stage. Generate an output state and pass it to the next moment





https://en.wikipedia.org/wiki/File:The_LSTM_cell.png

GAN

Do you know this two guys?

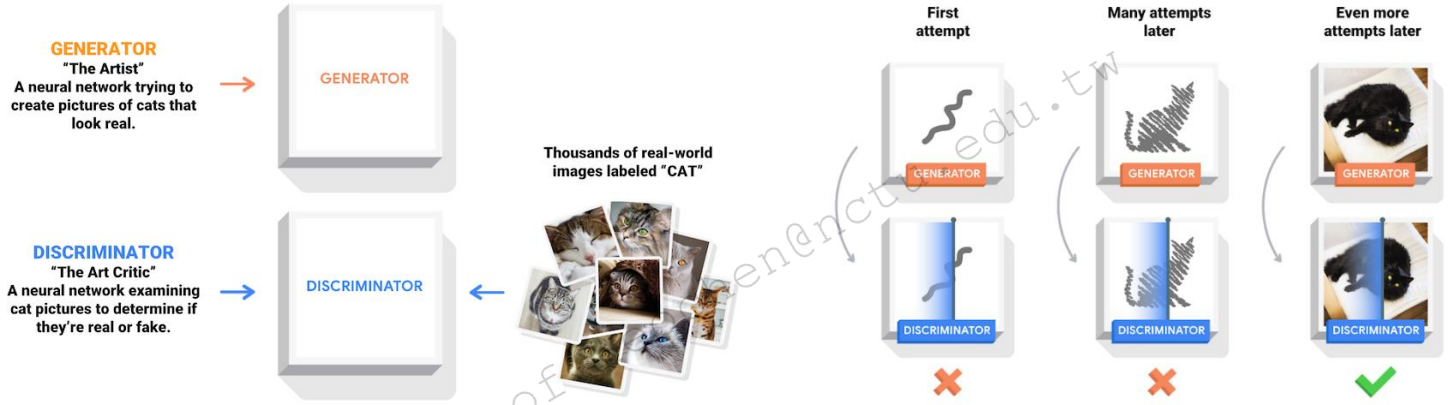
GAN

Do you know this two guys?

They are not real people in reality! They are generated by GAN!

GAN

- In 2016, Ian Goodfellow proposed Generative Adversarial Networks (GANs), which defines two networks: Generator (G) and Discriminator (D).
  - G is a generator that inputs a random noise and outputs a picture generated by random noise.
  - D is a discriminator that predicts if a picture is real or fake.
- The training process is to let the two models play against each other until the discriminator can hardly distinguish whether an image is real or generated. At this point, the generator can be used to "create" pictures that do not exist in reality.



https://www.tensorflow.org/beta/tutorials/generative/dcgan

Limitations of Deep Learning

About dataset:

1. Hungry for a large number of structural dataset, which is expensive in time, human resource and money.
2. Poor in few shots learning (only a small sample are given).

About model:

1. Focus on the understanding of unstructured data, instead of structured data.
2. The “learning” of deep learning is closer to memory, rather than reasoning.
3. Poor interpretability of deep neural networks.

About deploy:

- Generally speaking, the more powerful model, the more parameters, the larger computation. But we want to deploy it everywhere, including IOT devices.

Future of Deep Learning

Despite the tremendous achievements in deep learning, researchers have not stopped their exploration of the unknown.

Here we will introduce two new attempts in deep learning:

- Capsule networks
- Neural Architecture Search (NAS)

Qualcomm

Capsule networks

- Max-pooling with stride 2 is widely used in CNN. But Hinton criticized[1][2] that pooling layer discards positional information. (Convolution with stride 2 is similar.)

- The capsule network is proposed by Hinton et al., whose basic unit is a ***capsule***. Each capsule outputs a vector that contains *{likelihood, orientation, size}.*

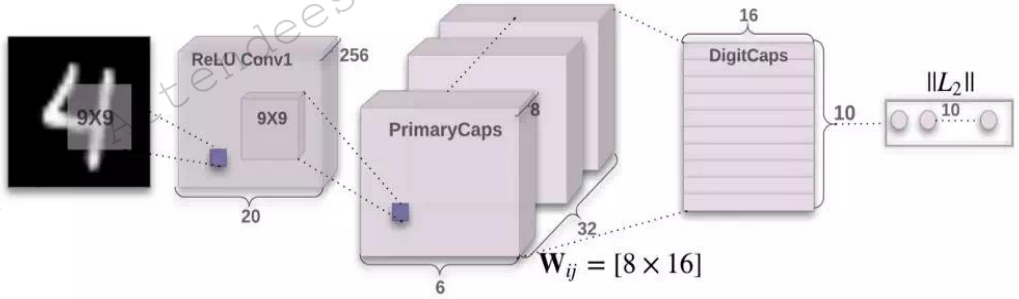
- If the objects in the area have spatial changes (shift, rotation, scaling, etc.), the capsule outputs a vector of the same length (representing the object existing or not), but in a different direction (representing spatial information of the object).

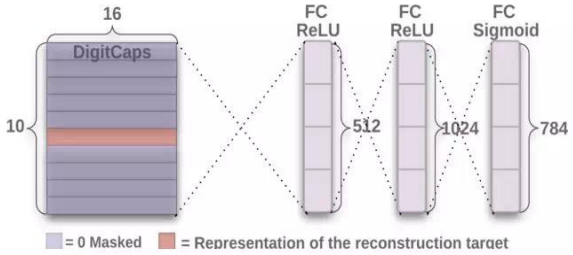
[1] https://www.reddit.com/r/MachineLearning/comments/2lmo0l/ama_geoffrey_hinton/clyj4jv/
[2] S Sabour, Dynamic Routing Between Capsules, arxiv, 1710.

Capsule networks

CapNet is divided into an encoder and a decoder.

- The encoder is used for classification. The first layer is the Convolution layer, and the last two layers are the Capsule layer, which calculates the classification loss.
- The decoder is used to restore the original picture, followed by Euclidean distance loss.



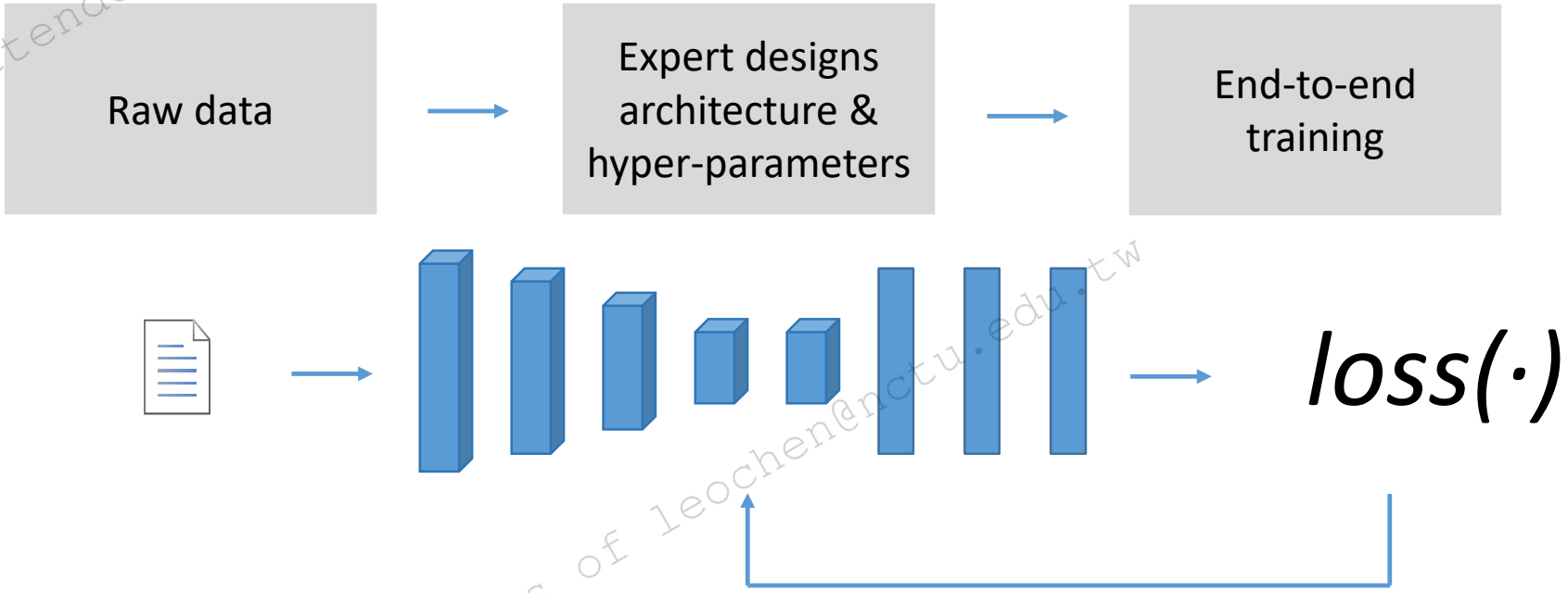


- The 3 layers encoder of CapNet reaches 0.25% error rate in MNIST, which is better than the baseline CNN model (0.39%).
- It should be pointed out that the capsule network is still in the research and experimental stage, and there are no other mature application scenarios.

Dynamic Routing Between Capsules, arxiv, 1710.09829

Neural Architecture Search (NAS)

- Current deep learning depends on experts to design network architecture and hyper-parameters.
- “In Deep Learning, Architecture Engineering is the New Feature Engineering.”



Neural Architecture Search (NAS)

- Neural Architecture Search (NAS) aims to jump out of the manual rules, and search **all possible architectures** by following a **search strategy**, then maximize the **performance metric** to find the optimal solution of network architecture.
- Networks by NAS may be a little messy and complicated, but they are can be more effective than expert-designed architecture.

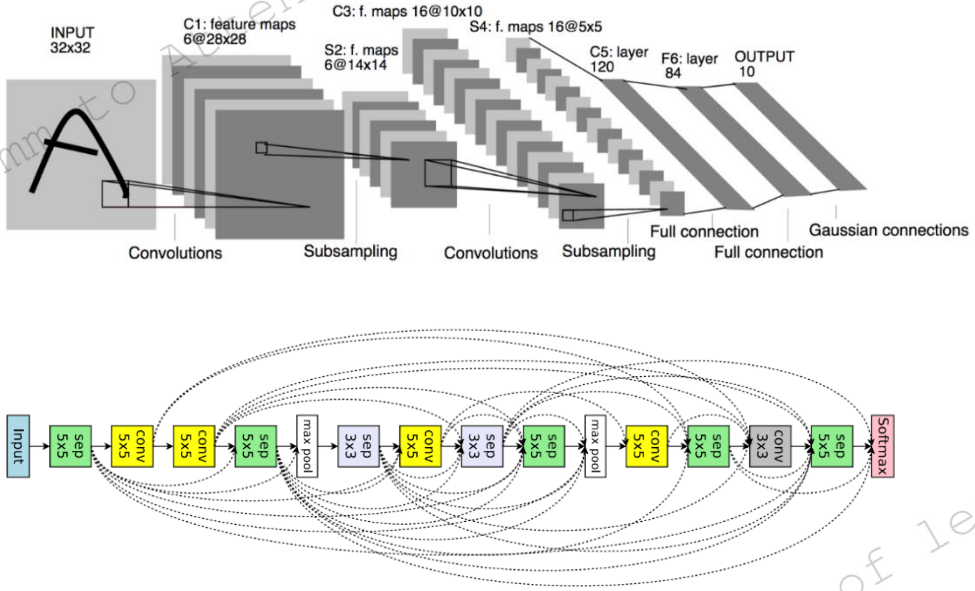


Figure 7. ENAS's discovered network from the macro search space for image classification.

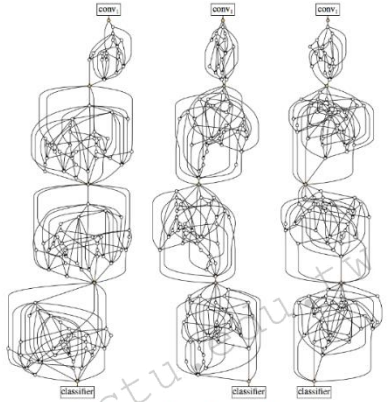


Figure 1. **Randomly wired neural networks** generated by the classical Watts-Strogatz (WS) [50] model: these three instances of random networks achieve (left-to-right) 79.1%, 79.1%, 79.0% classification accuracy on ImageNet under a similar computational budget to ResNet-50, which has 77.1% accuracy.

Gradient-based learning applied to document recognition, 1998
Efficient Neural Architecture Search via Parameter Sharing, arxiv 1802.03268
Exploring Randomly Wired Neural Networks for Image Recognition, arxiv 1904.01569
Neural Architecture Search: A Survey, arxiv 1808.05377

Thank You