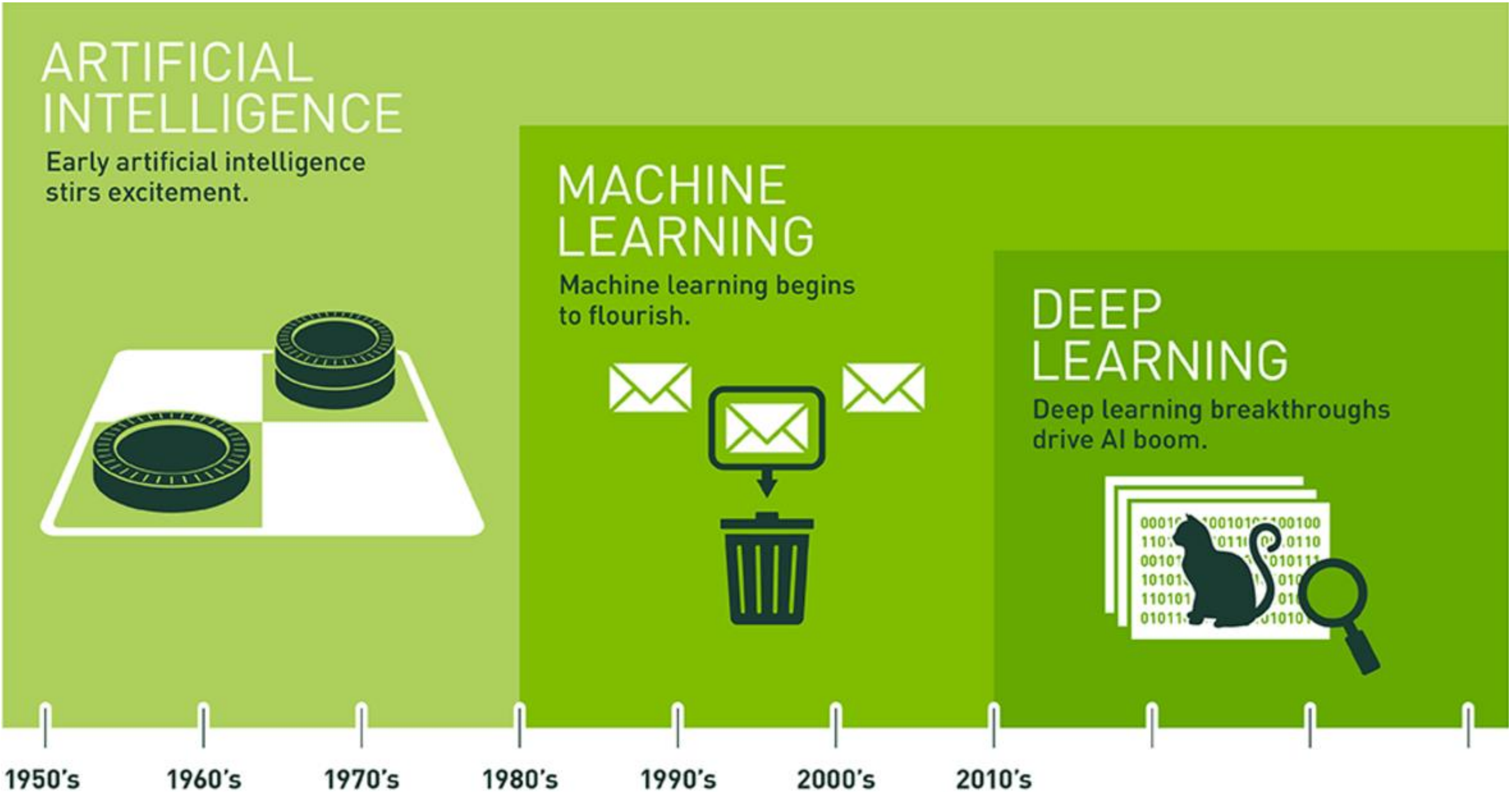
人工智慧 機器學習 深度學習 簡介

帶你回顧歷史 展望未來


ARTIFICIAL INTELLIGENCE
Early artificial intelligence stirs excitement.
MACHINE LEARNING
Machine learning begins to flourish.
DEEP LEARNING
Deep learning breakthroughs drive AI boom.
1950's
1960's
1970's
1980's
1990's
2000's
2010's

人工智慧 Artificial Intelligence（AI）

- 人工智慧:指由人製造出來的機器所表現出來的智慧。

人工智慧 Artificial Intelligence（AI）

- 人工智慧:指由人製造出來的機器所表現出來的智慧。
- 智慧是哲學問題

什麼是智慧

John Searle:「強人工智慧」(Strong A.I.)和

「弱人工智慧」(Weak A.I.)

強人工智慧:

強調電腦將能擁有自覺意識、性格、情感、知覺、社交等人類的特徵。

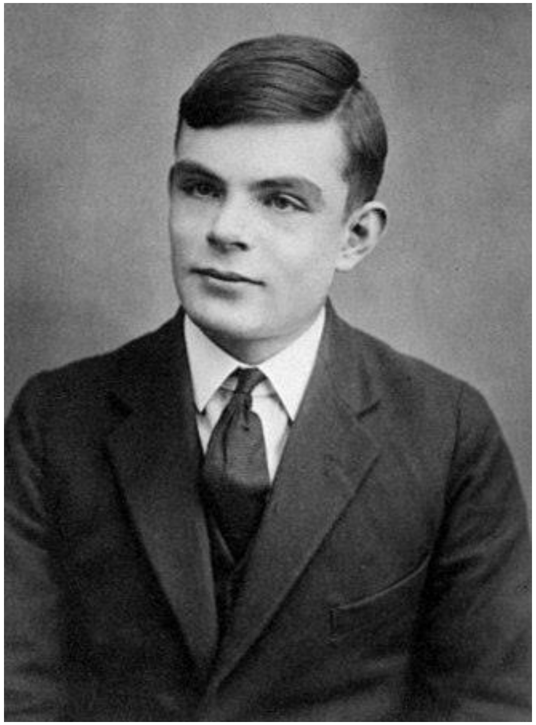
弱人工智慧:

主張機器只能模擬人類具有思維的行為表現，而不是真正懂得思考。他們認為機器僅能模擬人類，並不具意識、也不理解動作本身的意義。

Example:烏鴉跟鸚鵡

計算機科學與人工智慧之父

- 2014年的電影《模仿遊戲》
- 機器會思考嗎？
- 圖靈測試



人工智慧的方向

- 人工智慧的研究歷史有著一條從以「推理」為重點，到以「知識」為重點，再到以「學習」為重點的自然、清晰的脈絡。

黃金年代1956-1974

- 1955年，全世界第一個人工智慧程式之稱的邏輯理論家 (Logic Theorist)
- 1958年，十年之內，數字計算機將成為西洋棋世界冠軍。(1997 IBM深藍電腦打敗當時西洋棋世界冠軍)
- 1965年，二十年內，機器將能完成人能做到的一切工作。
- 1967年，一代之內，創造『人工智慧』的問題將獲得實質上的解決。
- 1970年，在三到八年的時間裡我們將得到一台具有人類平均智能的機器。

第一次AI低谷1974-1980

- 人類不知，機器也不能解決，AI這時沒超過人類。
- 對人工智慧的研究方向局限於邏輯數學領域、加上硬體環境上的困境，使早期人工智慧只能解一些代數題和數學證明，難以在實務上有所應用。

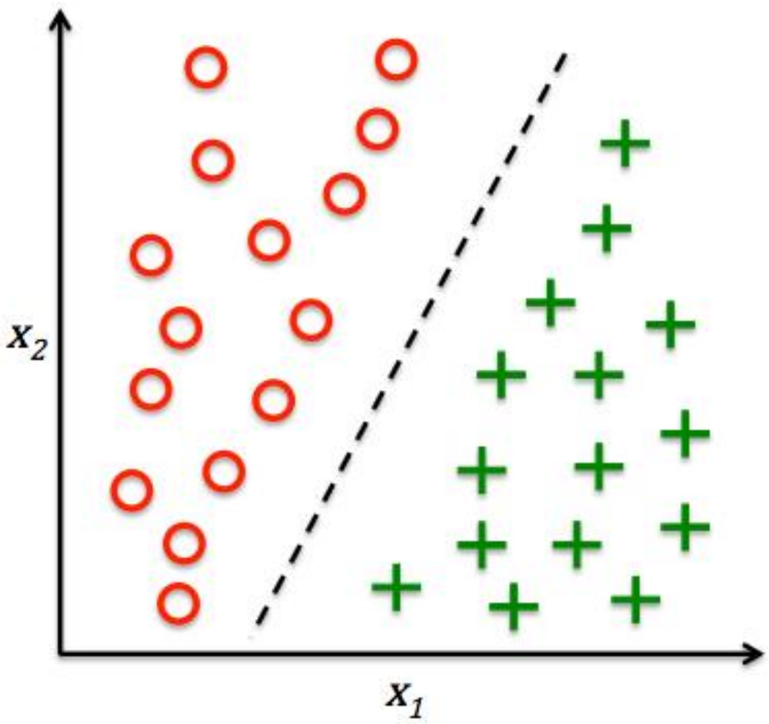
再次繁榮1980-1987

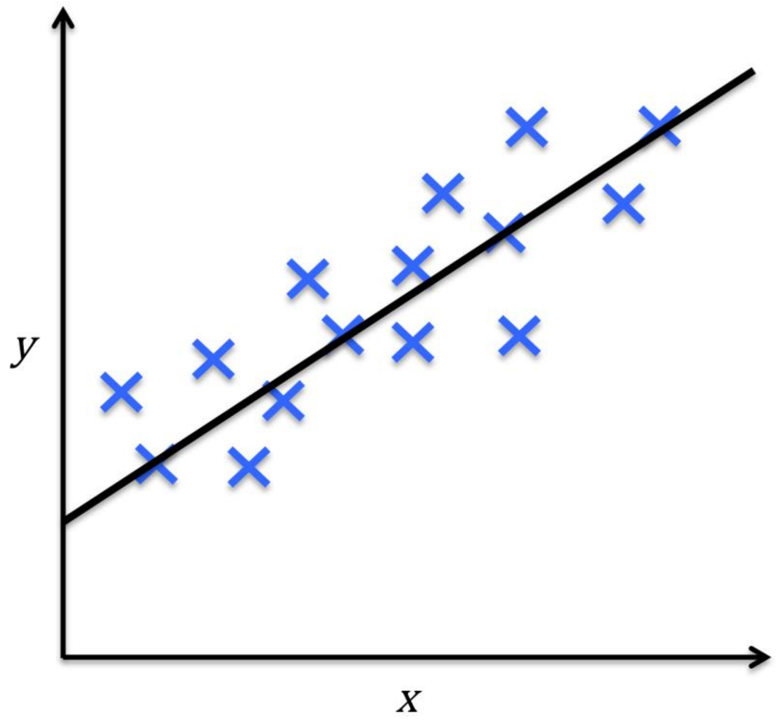
- 「機器學習」是一門涵蓋電腦科學、統計學、機率論、博弈論等多門領域的學科，從 1980 開始蓬勃興起。
- 第一次人工智慧泡沫後，研究領域轉為「機器學習」(Machine Learning)。
- 機器學習之所以能興起，也歸功於硬體儲存成本下降、運算能力增強 (包括本機端與雲端運算)，加上大量的數據能做處理。

機器學習的分類

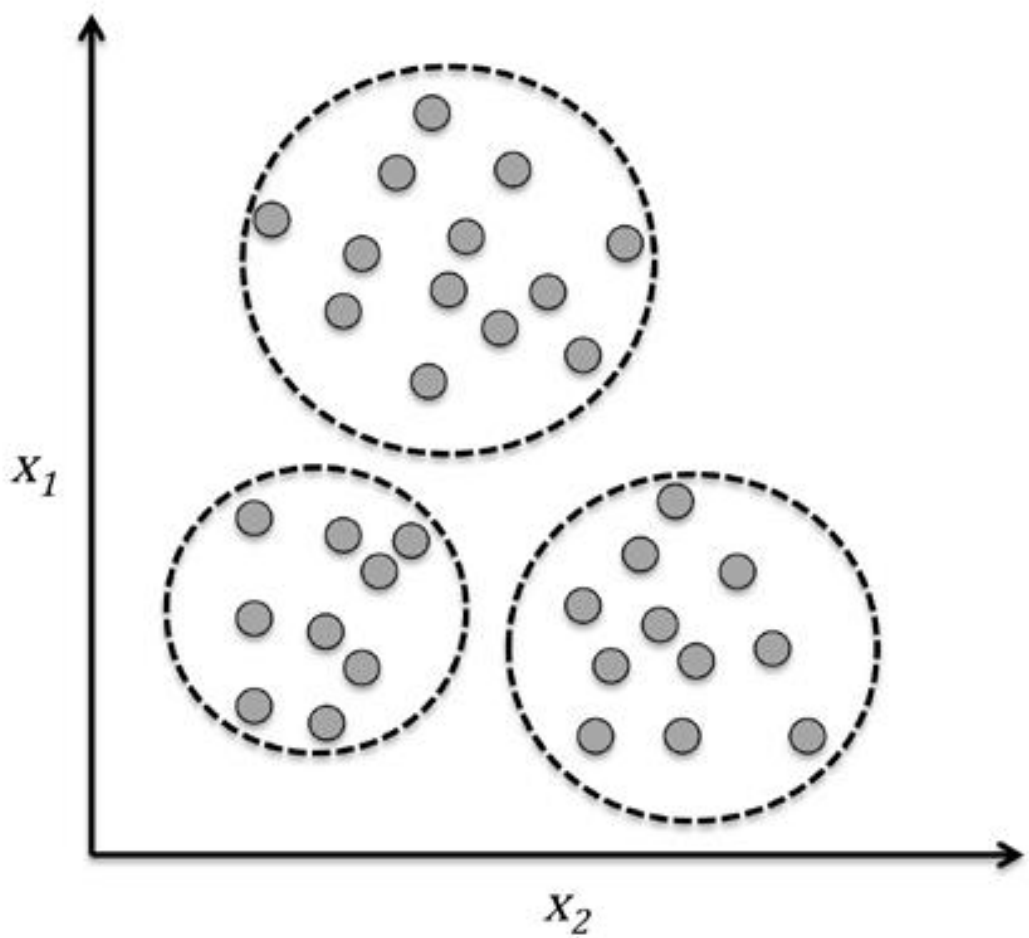
- 監督學習：通過已有的一部分輸入數據與輸出數據之間的對應關系，生成一個函數，將輸入映射到合適的輸出。
- 函數的輸出可以是一個連續的值（稱為迴歸分析），或是預測一個分類標籤（稱作分類）。
- 無監督學習：直接對輸入數據集進行建模，例如聚類。
- 強化學習：一面觀察，一面學習。

監督學習

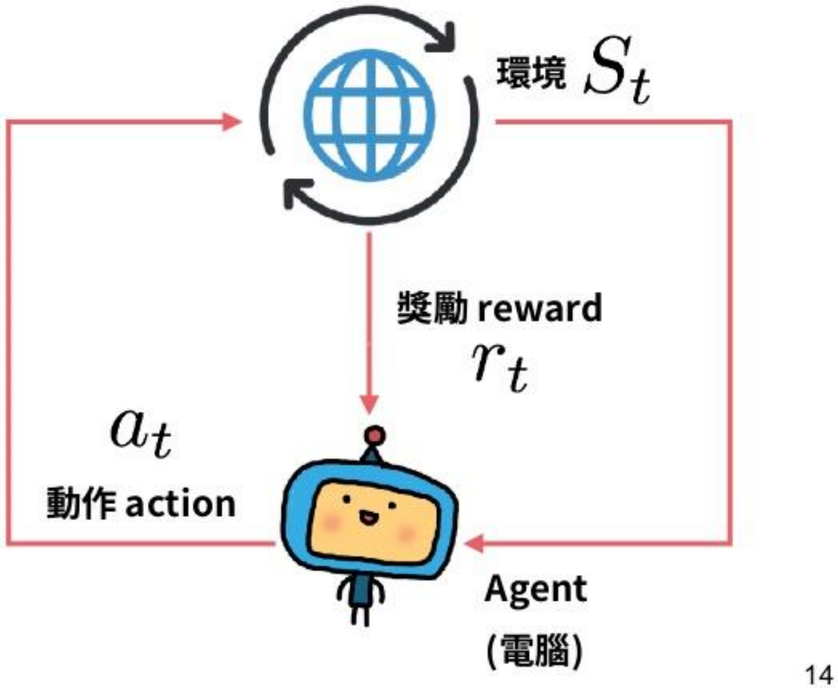




無監督學習
分群問題

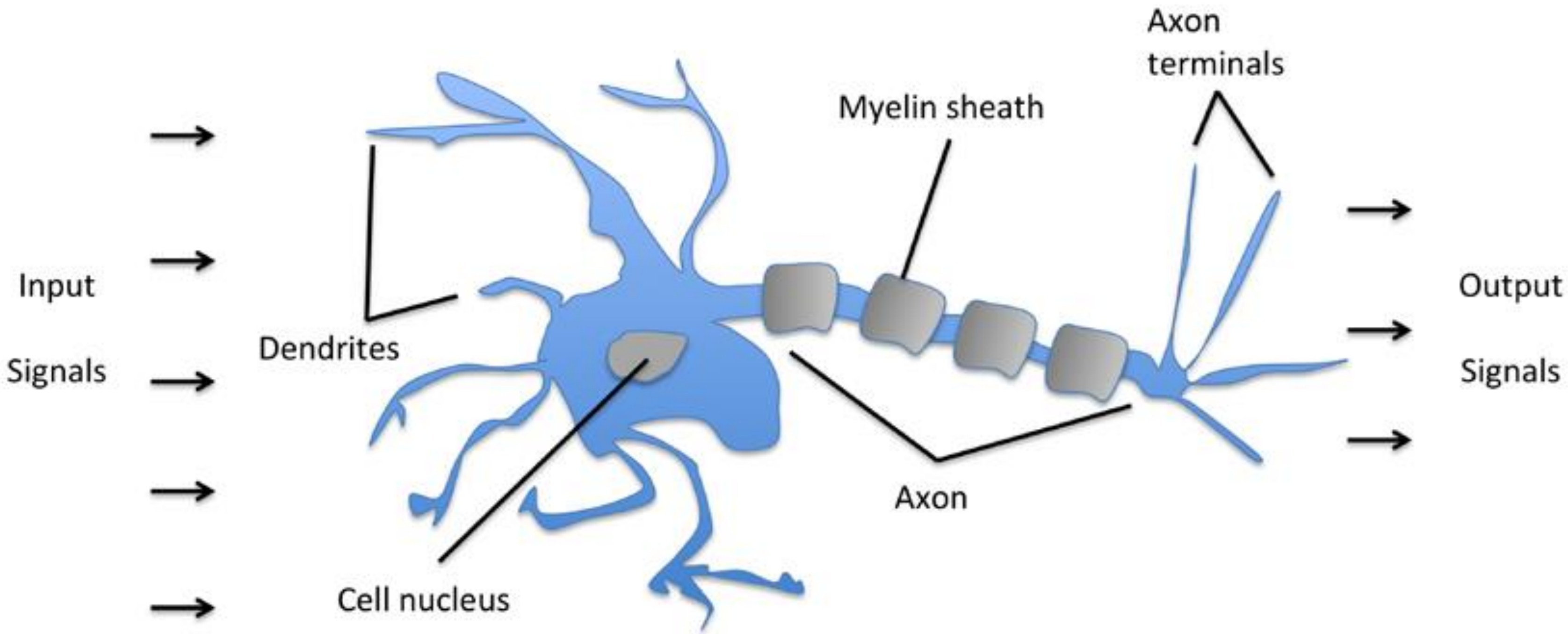


強化學習(下棋)

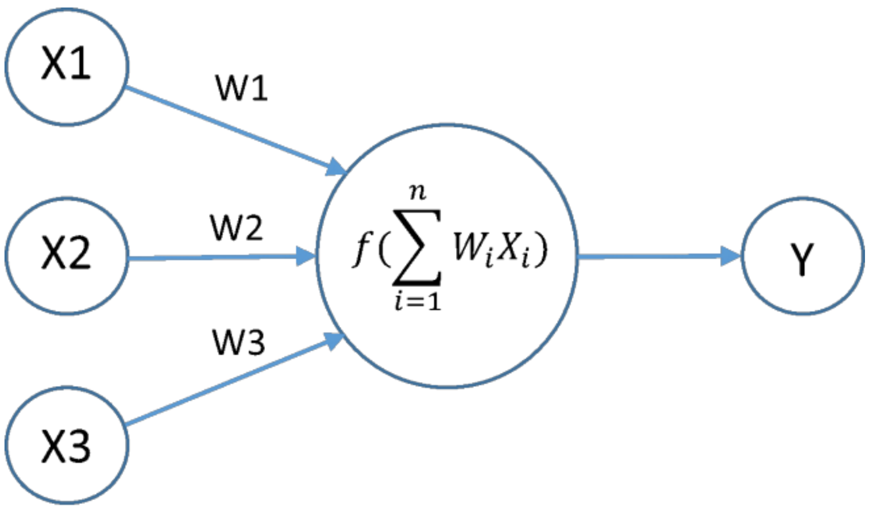


類神經網路

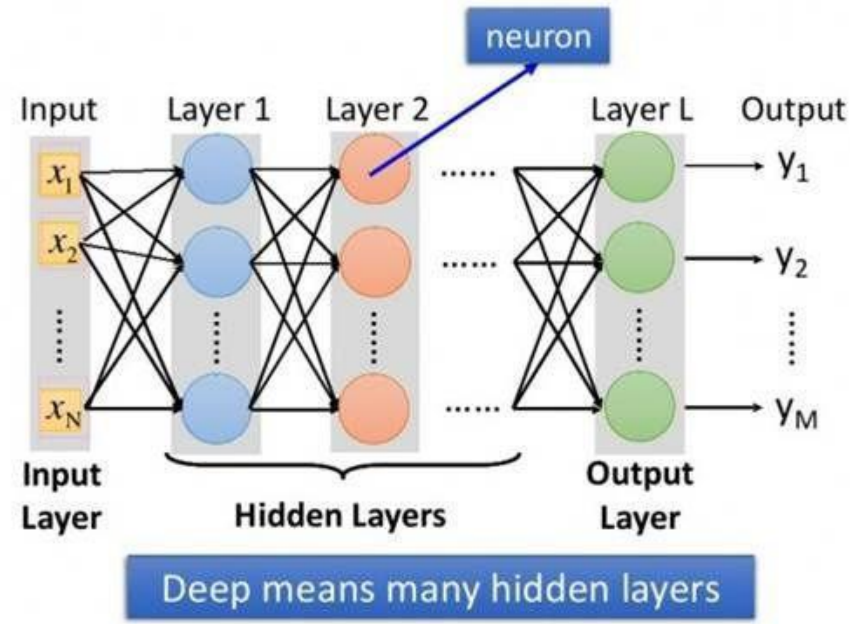
- 1943年提出「MCP神經元」用來描述一個簡化的腦細胞。



類神經網路

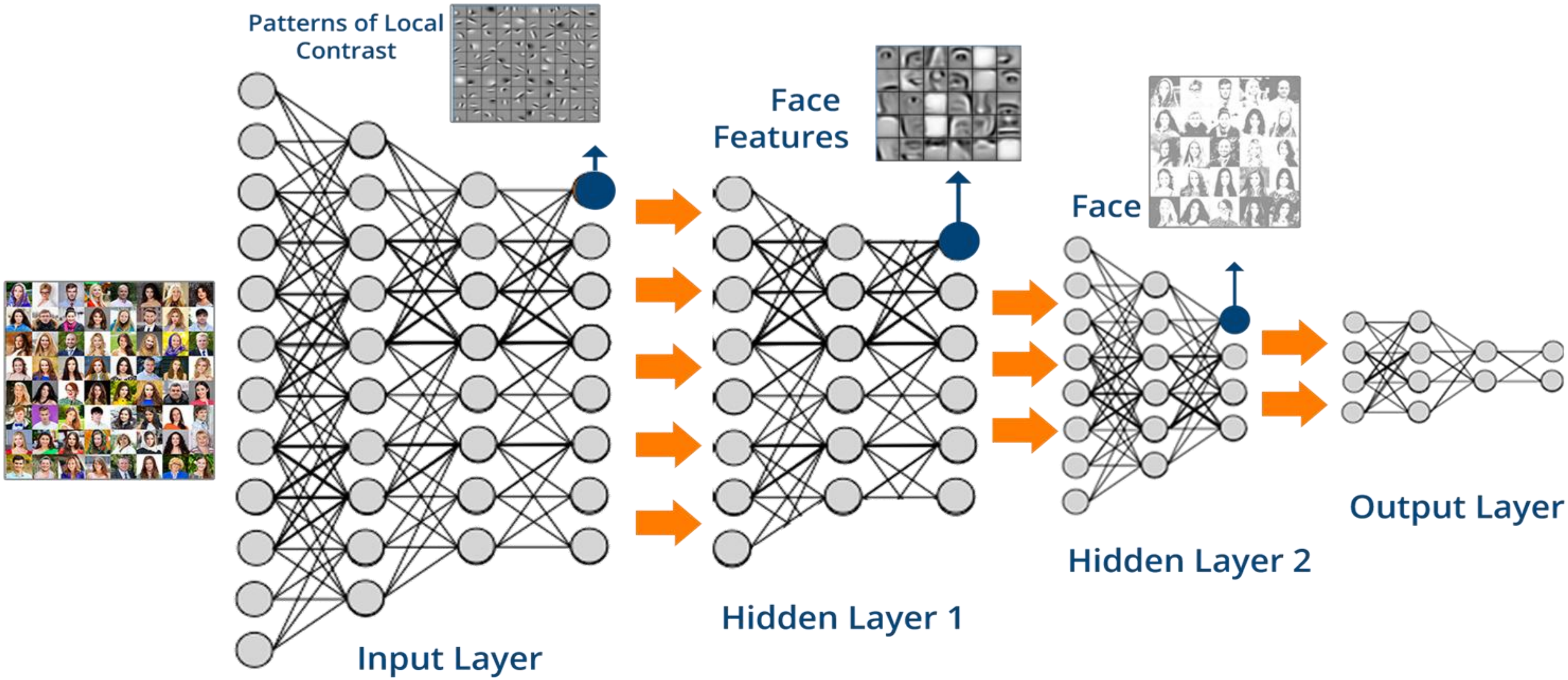


X1
X2
X3
W1
W2
W3
$f(\sum_{i=1}^{n} W_i X_i)$
Y

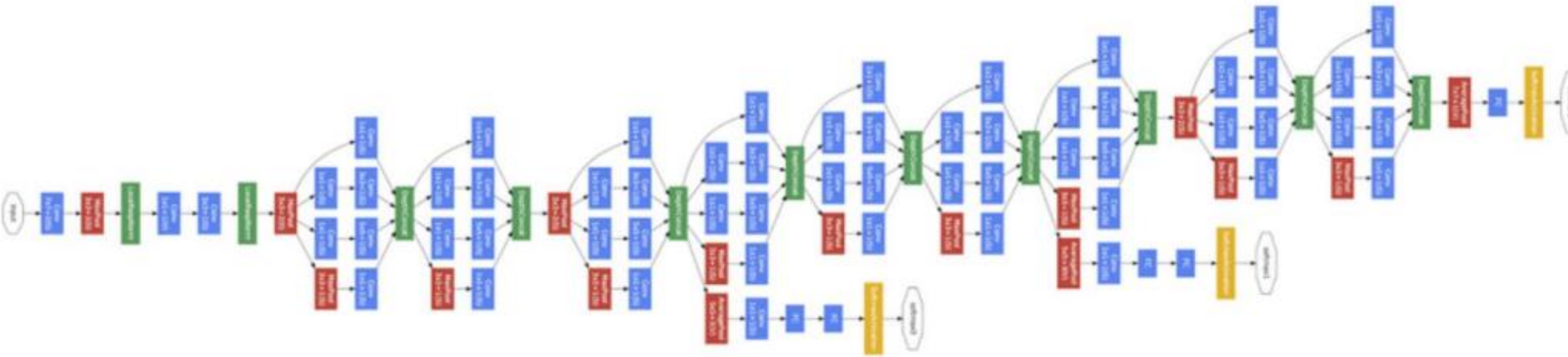


neuron
Input
Layer 1
Layer 2
Layer L
Output
x_1
x_2
x_N
y_1
y_2
y_M
Input Layer
Hidden Layers
Output Layer
Deep means many hidden layers

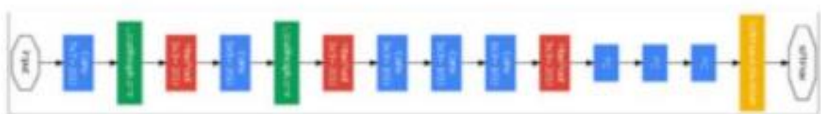
CNN


Patterns of Local Contrast
Face Features
Face
Input Layer
Hidden Layer 1
Hidden Layer 2
Output Layer

GoogLeNet vs AlexNet



GoogLeNet



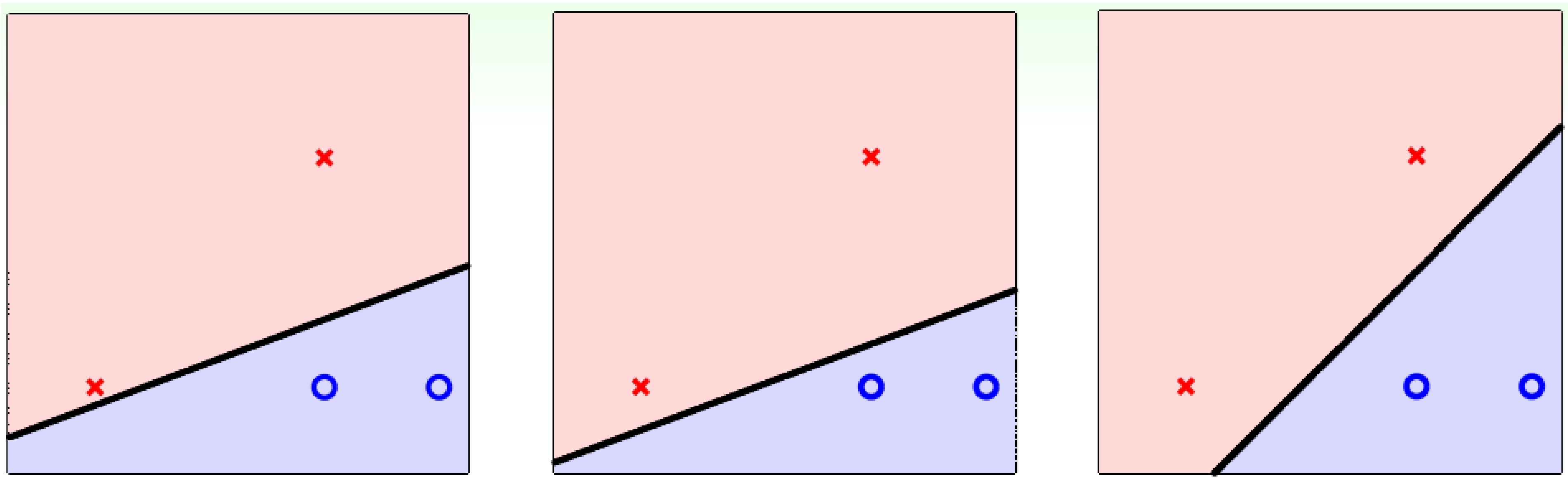
AlexNet

Convolution
Pooling
Softmax
Other

深度學習的低谷

- 淺層的機器學習模型。
- GPU沒出來，運算能力不行，數據不夠多。
- 1990年，支撐向量機 (SVM, Support Vector Machines)廣受歡迎。

SVM簡介

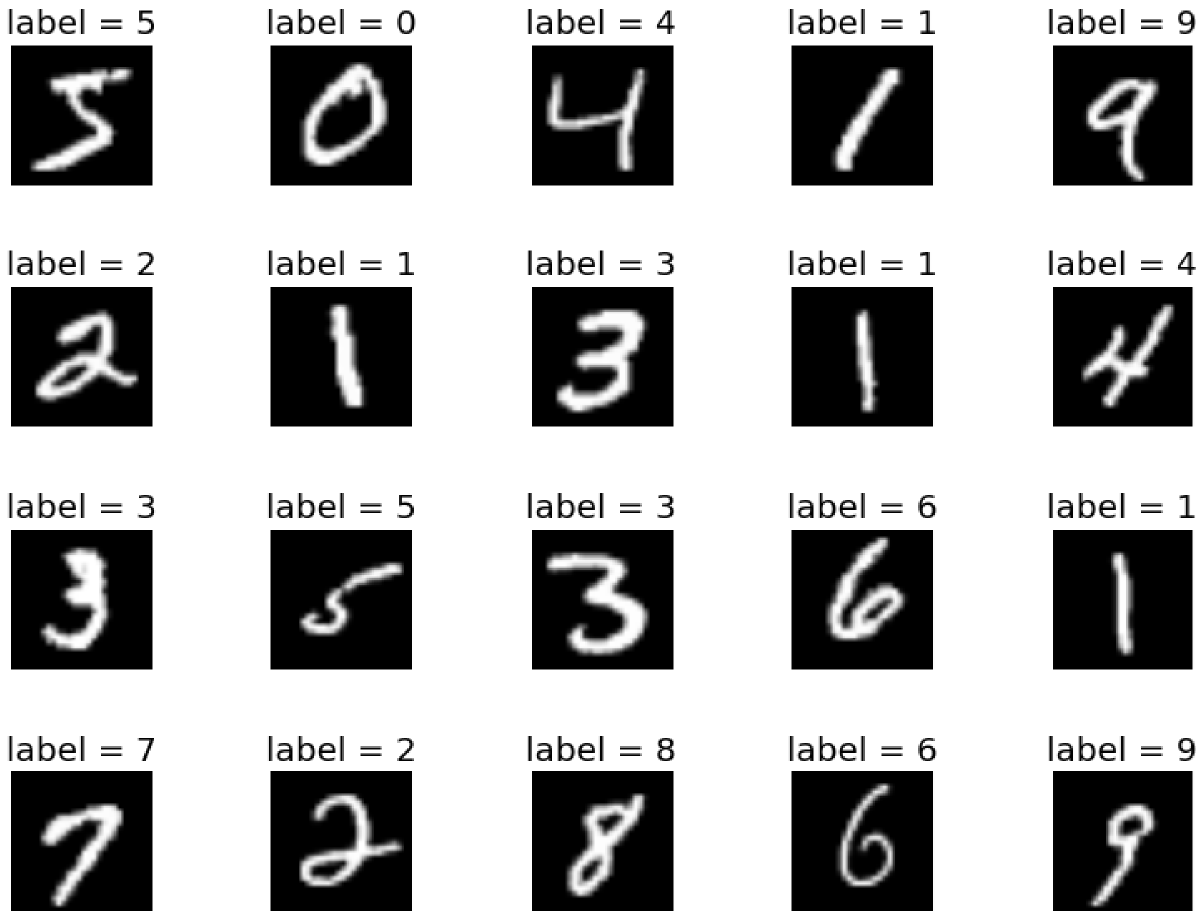


2012再次爆發人工智慧的熱潮

- ImageNet 是全世界最大的圖像識別資料庫。每年，史丹佛大學都會舉辦 ImageNet 圖像識別競賽，參加者包括了Google、微軟、百度等大型企業，除了在比賽中爭奪圖像識別寶座、同時測試自家系統的效能與極限。
- 其實從 2007 年 ImageNet 比賽創辦以來，每年的比賽結果、每家都差不多，錯誤率大致落在 30%、29%、28%… 瓶頸一直無法突破。
- 結果 2012 年 Hinton 的兩個學生以 SuperVision(AlexNet) 的隊伍名參賽，以 16.42% 的錯誤率遠勝第二名的 26.22%。用的正是深度學習技術。

分享學習會碰到的問題

最常用的入門應用


label = 5
label = 0
label = 4
label = 1
label = 9
label = 2
label = 1
label = 3
label = 1
label = 4
label = 3
label = 5
label = 3
label = 6
label = 1
label = 7
label = 2
label = 8
label = 6
label = 9

自動生成文章
Blue score, Perplexity

Abstract
Learning to store information over extended time intervals via recurrent backpropagation takes a very long time,mostly due to insuﬃcient,decaying error back ♪ow.We brie♪y review Hochreiter's 1991analysis of this problem,then address it by introducing a novel,eﬃcient, gradient-based method called ₩Long Short-Term Memory"(LSTM).Truncating the gradient where this does not do harm,LSTM can learn to bridge minimal time lags in excess of 1000 discrete time steps by enforcing constanterror ♪ow through ₩constant error carrousels"within special units.Multiplicative gate units learn to open and close access to the constant error ♪ow.LSTM is local in space and time;its computational complexity per time step and weight is O(1).Our experiments with arti♀cial data involve local,distributed,real-valued,and noisy pattern representations.In comparisons with RTRL,BPTT,Recurrent Cascade-Correlation, Elman nets,and Neural Sequence Chunking,LSTM leads to many more successful runs,and learns much faster.LSTM also solves complex,arti♀cial long time lag tasks that have never been solved by previous recurrent network algorithms.

1INTRODUCTION

Recurrent networks can in principle use their feedback connections to store representations of recent input events in form of activations (₩short-term memory",as opposed to ₩long-term memory"embodied by slowly changing weights).This is potentially signi♀cant for many applications, including speech processing,non-Markovian control,and music composition (e.g.,Mozer 1992). The most widely used algorithms for learning what to put in short-term memory,however,take too much time or do not work well at all,especially when minimal time lags between inputs and corresponding teacher signals are long.Although theoretically fascinating,existing methods do not provide clear practical advantages over,say,backprop in feedforward nets with limited time windows.This paper will review an analysis of the problem and suggest a remedy.

The problem.With conventional ₩Back-Propagation Through Time"(BPTT,e.g.,Williams and Zipser 1992,Werbos 1988)or ₩Real-Time Recurrent Learning"(RTRL,e.g.,Robinson and Fallside 1987),error signals ₩♪owing backwards in time"tend to either (1)blow up or (2)vanish: the temporal evolution of the backpropagated error exponentially depends on the size of the weights (Hochreiter 1991).Case (1)may lead to oscillating weights,while in case (2)learning to bridge long time lags takes a prohibitive amount of time,or does not work at all (see section 3).

The remedy.This paper presents ₩Long Short-Term Memory"(LSTM),a novel recurrent network architecture in conjunction with an appropriate gradient-based learning algorithm.LSTM is designed to overcome these error back-♪ow problems.It can learn to bridge time intervals in excess of 1000steps even in case of noisy,incompressible input sequences,without loss of short time lag capabilities.This is achieved by an eﬃcient,gradient-based algorithm for an architecture 1enforcing constant (thus neither exploding nor vanishing)error ♪ow through internal states of special units (provided the gradient computation is truncated at certain architecture-speci♀c points |this does not a♂ect long-term error ♪ow though).

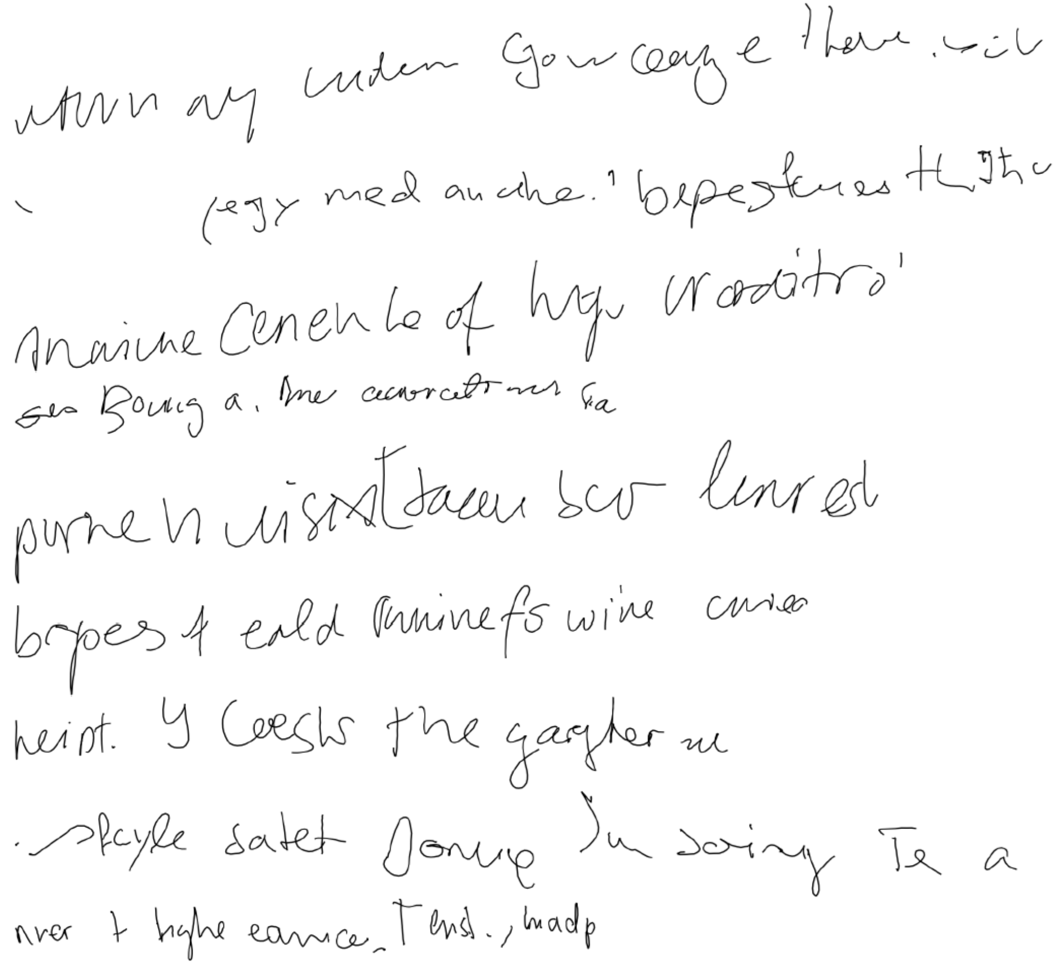
Outline of paper.Section 2will brie♪y review previous work.Section 3begins with an outline of the detailed analysis of vanishing errors due to Hochreiter (1991).It will then introduce a naive approach to constant error backprop for didactic purposes,and highlight its problems concerning information storage and retrieval.These problems will lead to the LSTM architecture as described in Section 4.Section 5will present numerous experiments and comparisons with competing methods.LSTM outperforms them,and also learns to solve complex,arti♀cial tasks no other recurrent net algorithm has solved.Section 6will discuss LSTM's limitations and advantages.The appendix contains a detailed description of the algorithm (A.1),and explicit error ♪ow formulae (A.2).

2PREVIOUS WORK

This section will focus on recurrent nets with time-varying inputs (as opposed to nets with stationary inputs and ♀xpoint-based gradient calculations,e.g.,Almeida 1987,Pineda 1987).
Gradient-descent variants.The approaches of Elman (1988),Fahlman (1991),Williams (1989),Schmidhuber (1992a),Pearlmutter (1989),and many of the related algorithms in Pearlmutter's

自動生成音樂
Music Generation

自動生成手寫字



For $\bigoplus_{n=1,\dots,m}$ where $\mathcal{L}_{m_\bullet} = 0$, hence we can find a closed subset $\mathcal{H}$ in $\mathcal{H}$ and any sets $\mathcal{F}$ on X, U is a closed immersion of S, then $U \to T$ is a separated algebraic space.

Proof. Proof of (1). It also start we get

$$S = \text{Spec}(R) = U \times_X U \times_X U$$

and the comparicoly in the fibre product covering we have to prove the lemma generated by $\coprod Z \times_U U \to V$. Consider the maps M along the set of points Sch_{fppf} and $U \to U$ is the fibre category of S in U in Section, **??** and the fact that any U affine, see Morphisms, Lemma **??**. Hence we obtain a scheme S and any open subset $W \subset U$ in $Sh(G)$ such that $\text{Spec}(R') \to S$ is smooth or an

$$U = \bigcup U_i \times_{S_i} U_i$$

which has a nonzero morphism we may assume that f_i is of finite presentation over S. We claim that $\mathcal{O}_{X,x}$ is a scheme where $x, x', s'' \in S'$ such that $\mathcal{O}_{X,x'} \to \mathcal{O}'_{X',x'}$ is separated. By Algebra, Lemma **??** we can define a map of complexes $\text{GL}_{S'}(x'/S'')$ and we win. □

To prove study we see that $\mathcal{F}|_U$ is a covering of $\mathcal{X}'$, and $\mathcal{T}_i$ is an object of $\mathcal{F}_{X/S}$ for $i > 0$ and $\mathcal{F}_p$ exists and let $\mathcal{F}_i$ be a presheaf of $\mathcal{O}_X$-modules on $\mathcal{C}$ as a $\mathcal{F}$-module. In particular $\mathcal{F} = U/\mathcal{F}$ we have to show that

$$\widetilde{M}^\bullet = \mathcal{I}^\bullet \otimes_{\text{Spec}(k)} \mathcal{O}_{S,s} - i_X^{-1}\mathcal{F})$$

is a unique morphism of algebraic stacks. Note that

$$\text{Arrows} = (Sch/S)^{opp}_{fppf}, (Sch/S)_{fppf}$$

and

$$V = \Gamma(S, \mathcal{O}) \longmapsto (U, \text{Spec}(A))$$

is an open subset of X. Thus U is affine. This is a continuous map of X is the inverse, the groupoid scheme S.

Proof. See discussion of sheaves of sets. □

The result for prove any open covering follows from the less of Example **??**. It may replace S by $X_{spaces,étale}$ which gives an open subspace of X and T equal to S_{Zar}, see Descent, Lemma **??**. Namely, by Lemma **??** we see that R is geometrically regular over S.

Lemma 0.1. *Assume (3) and (3) by the construction in the description.*

Suppose $X = \lim |X|$ (by the formal open covering X and a single map $\underline{Proj}_X(\mathcal{A}) = \text{Spec}(B)$ over U compatible with the complex

$$Set(\mathcal{A}) = \Gamma(X, \mathcal{O}_{X,\mathcal{O}_X}).$$

*When in this case of to show that $\mathcal{Q} \to \mathcal{C}_{Z/X}$ is stable under the following result in the second conditions of (1), and (3). This finishes the proof. By Definition **??** (without element is when the closed subschemes are catenary. If T is surjective we may assume that T is connected with residue fields of S. Moreover there exists a closed subspace $Z \subset X$ of X where U in X' is proper (some defining as a closed subset of the uniqueness it suffices to check the fact that the following theorem*

(1) *f is locally of finite type. Since $S = \text{Spec}(R)$ and $Y = \text{Spec}(R)$.*

Proof. This is form all sheaves of sheaves on X. But given a scheme U and a surjective étale morphism $U \to X$. Let $U \cap U = \coprod_{i=1,\dots,n} U_i$ be the scheme X over S at the schemes $X_i \to X$ and $U = \lim_i X_i$. □

The following lemma surjective restrocomposes of this implies that $\mathcal{F}_{x_0} = \mathcal{F}_{x_0} = \mathcal{F}_{X,\dots,0}$.

Lemma 0.2. *Let X be a locally Noetherian scheme over S, $E = \mathcal{F}_{X/S}$. Set $\mathcal{I} = \mathcal{J}_1 \subset \mathcal{I}'_n$. Since $\mathcal{I}^n \subset \mathcal{I}^n$ are nonzero over $i_0 \leq \mathfrak{p}$ is a subset of $\mathcal{J}_{n,0} \circ \overline{A}_2$ works.*

Lemma 0.3. *In Situation **??**. Hence we may assume $\mathfrak{q}' = 0$.*

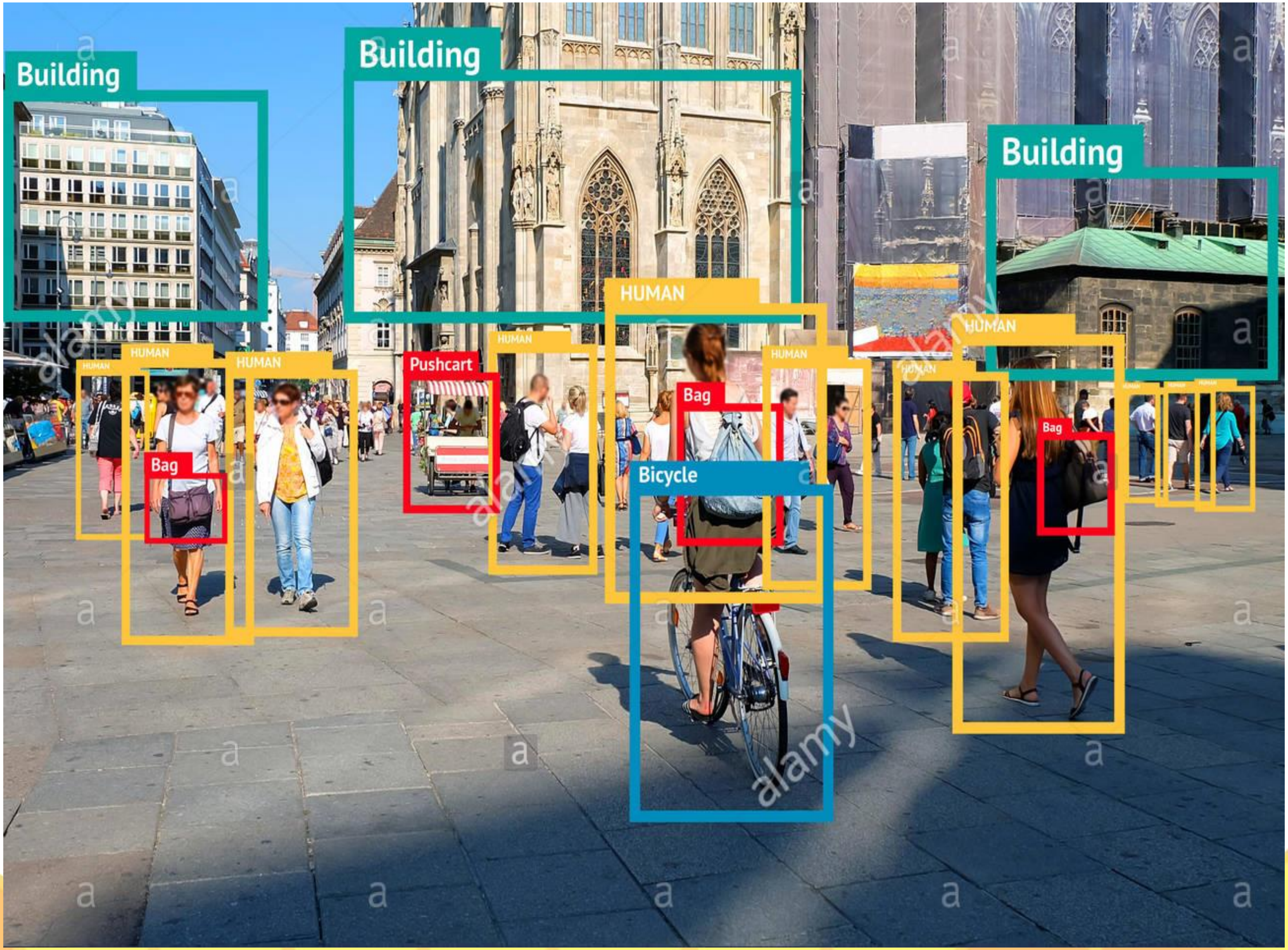
Proof. We will use the property we see that $\mathfrak{p}$ is the mext functor (**??**). On the other hand, by Lemma **??** we see that

$$D(\mathcal{O}_{X'}) = \mathcal{O}_X(D)$$

where K is an F-algebra where δ_{n+1} is a scheme over S. □

讓電腦自己玩遊戲

物體辨識


Building
Building
Building
HUMAN
HUMAN
HUMAN
HUMAN
HUMAN
HUMAN
HUMAN
HUMAN
Pushcart
Bag
Bag
Bag
Bicycle
alamy

分享深度學習的一些有趣研究
Style Transfer

1 **Upload photo**

The first picture defines the scene you would like to have painted.

2 **Choose style**

Choose among predefined styles or upload your own style image.

3 **Submit**

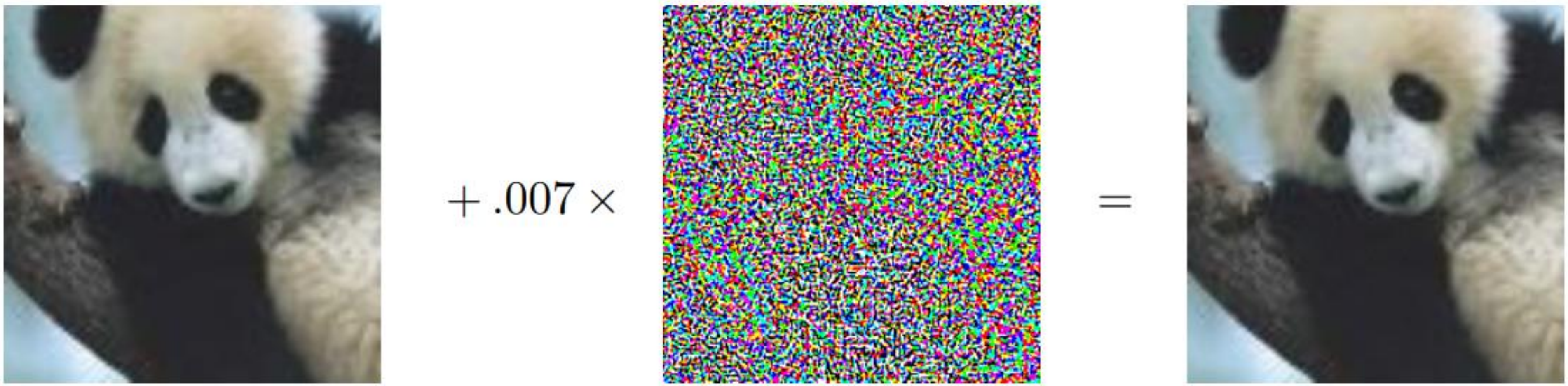
Our servers paint the image for you. You get an email when it's done.

分享深度學習的一些有趣研究

分享深度學習的一些有趣研究
Deep Dream

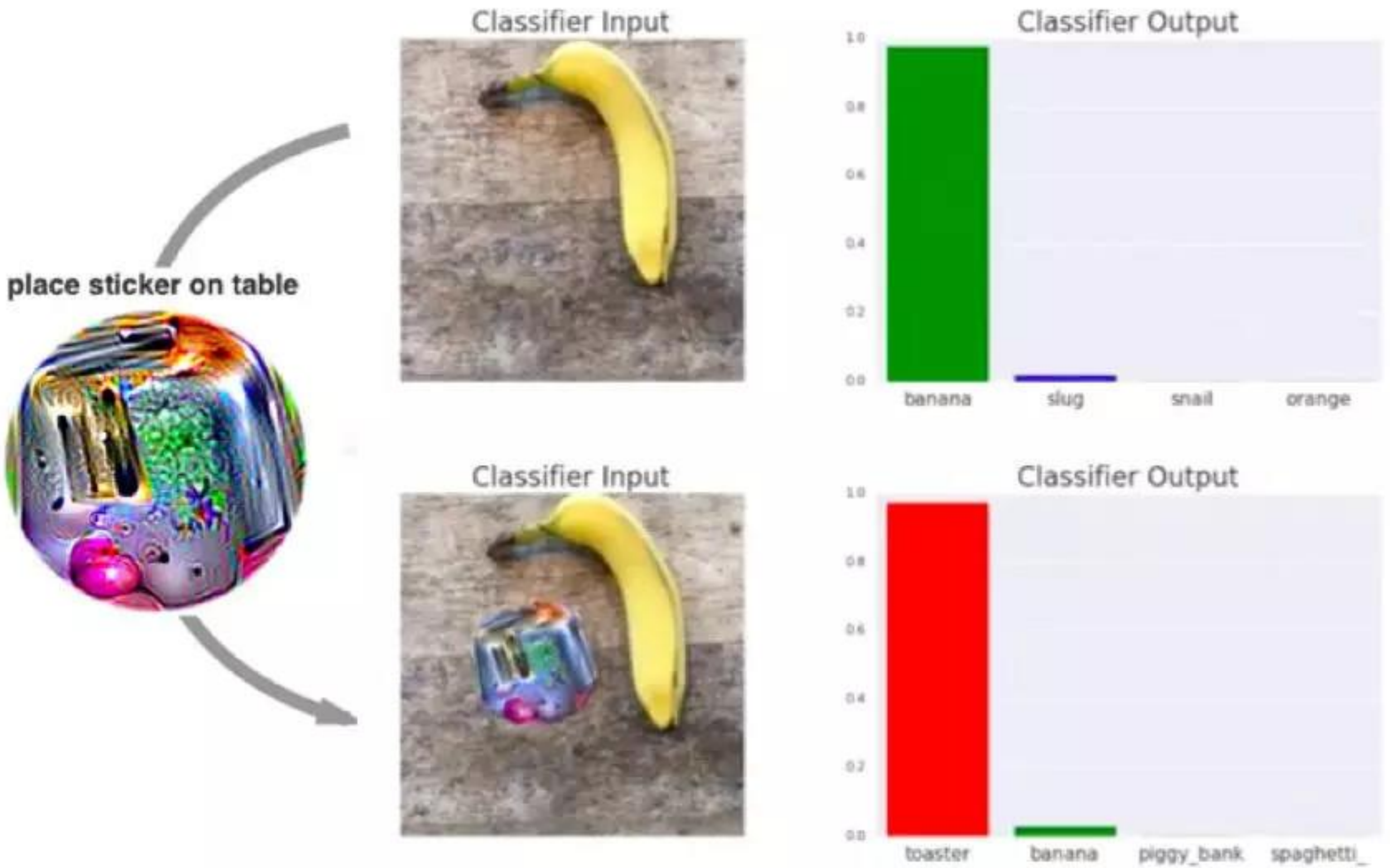
分享深度學習的一些有趣研究

深度學習結構的反思



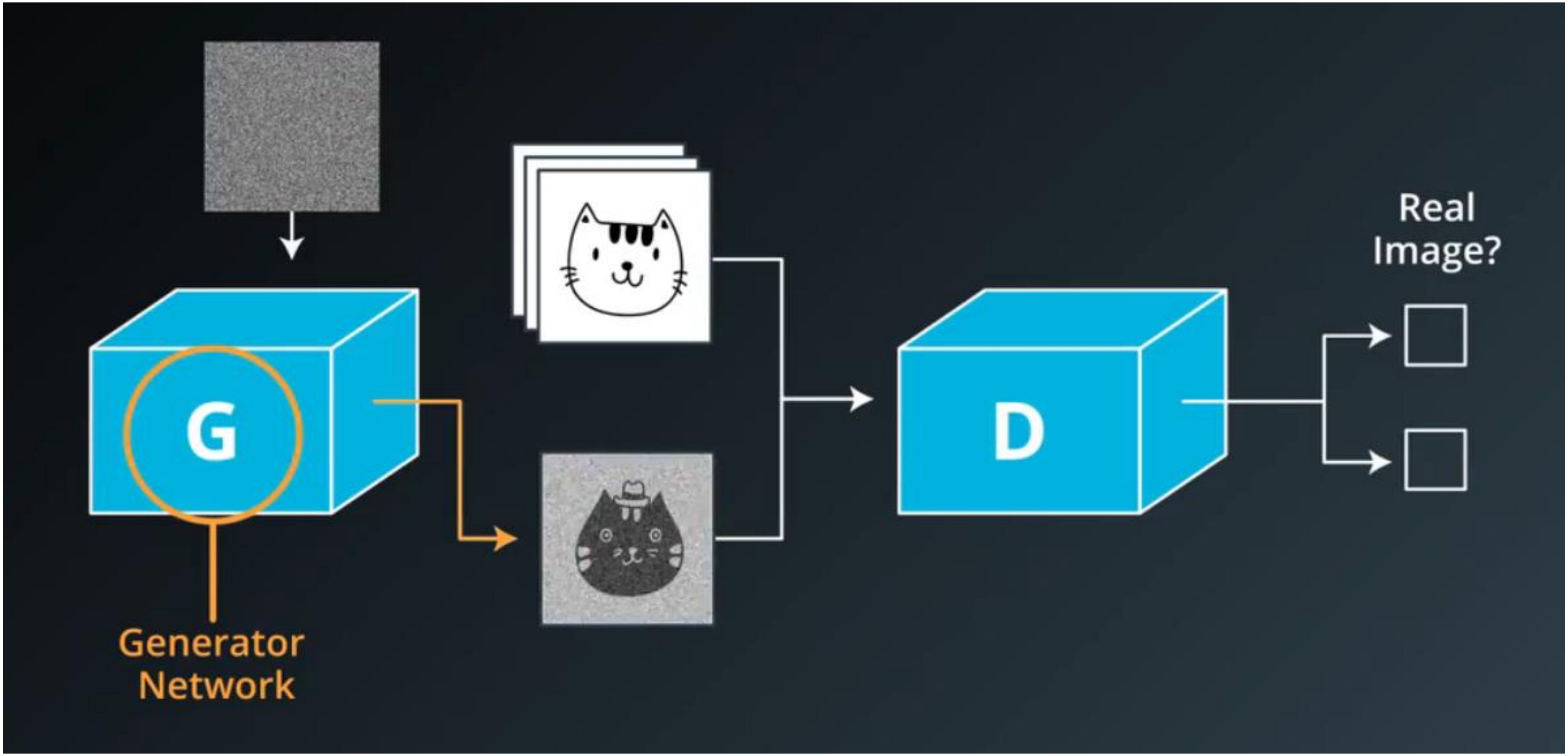
x "panda" 57.7% confidence $+ .007 \times$ $\text{sign}(\nabla_{x} J(\theta, x, y))$ "nematode" 8.2% confidence $=$ $x + \epsilon\text{sign}(\nabla_{x} J(\theta, x, y))$ "gibbon" 99.3 % confidence

深度學習結構的反思



place sticker on table
Classifier Input
Classifier Output
1.0
0.8
0.6
0.4
0.2
0.0
banana
slug
snail
orange
Classifier Input
Classifier Output
1.0
0.8
0.6
0.4
0.2
0.0
toaster
banana
piggy_bank
spaghetti_

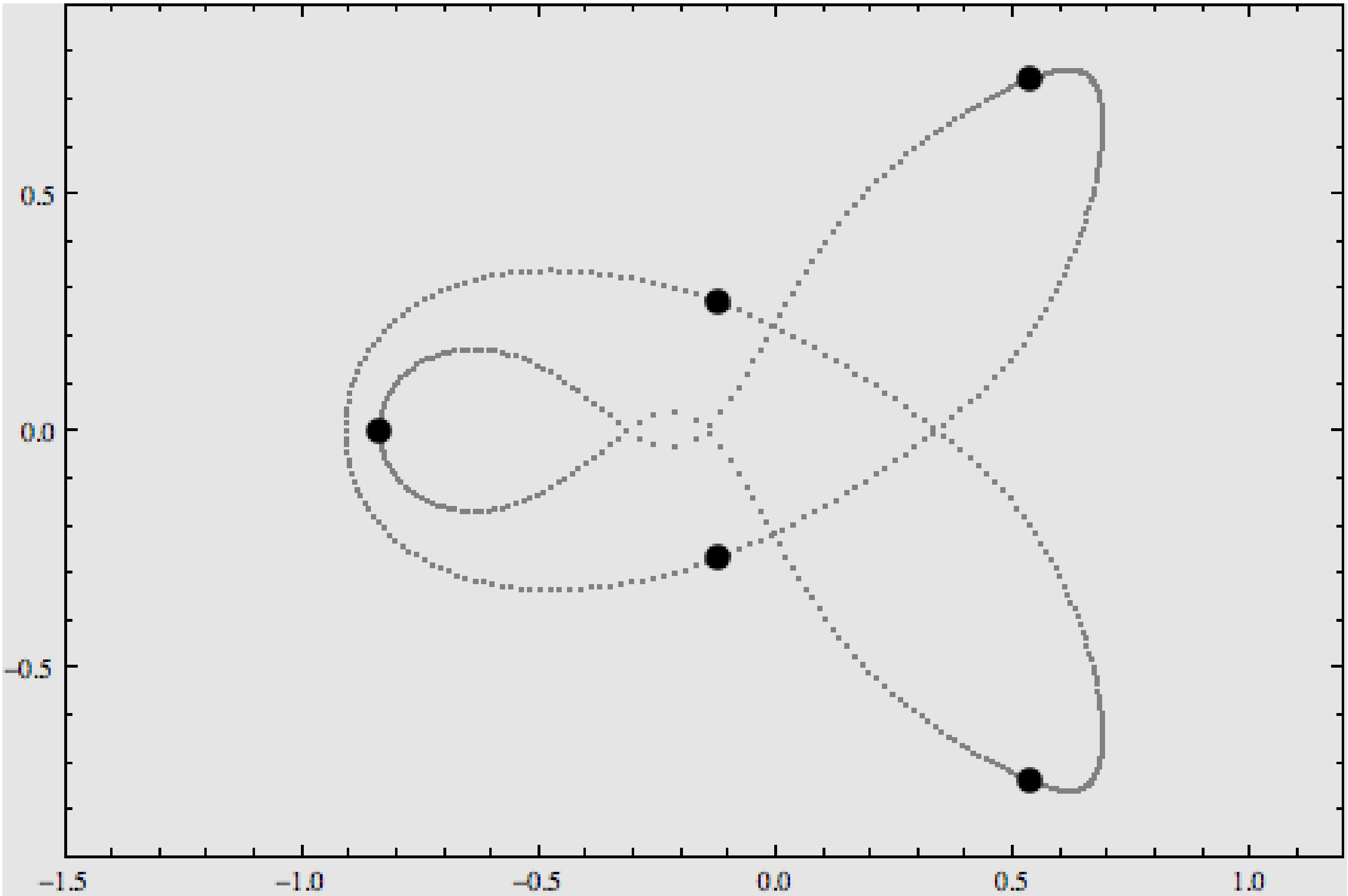
生成對抗網路
Generative Adversarial Network(GAN)



賽局理論Game Theory

	甲不坦白	甲坦白
乙不坦白	甲服刑1年 乙服刑1年	甲无罪释放 乙服刑15年
乙坦白	甲服刑15年 乙无罪释放	甲服刑6年 乙服刑6年

多體運動週期解



自動化機器學習

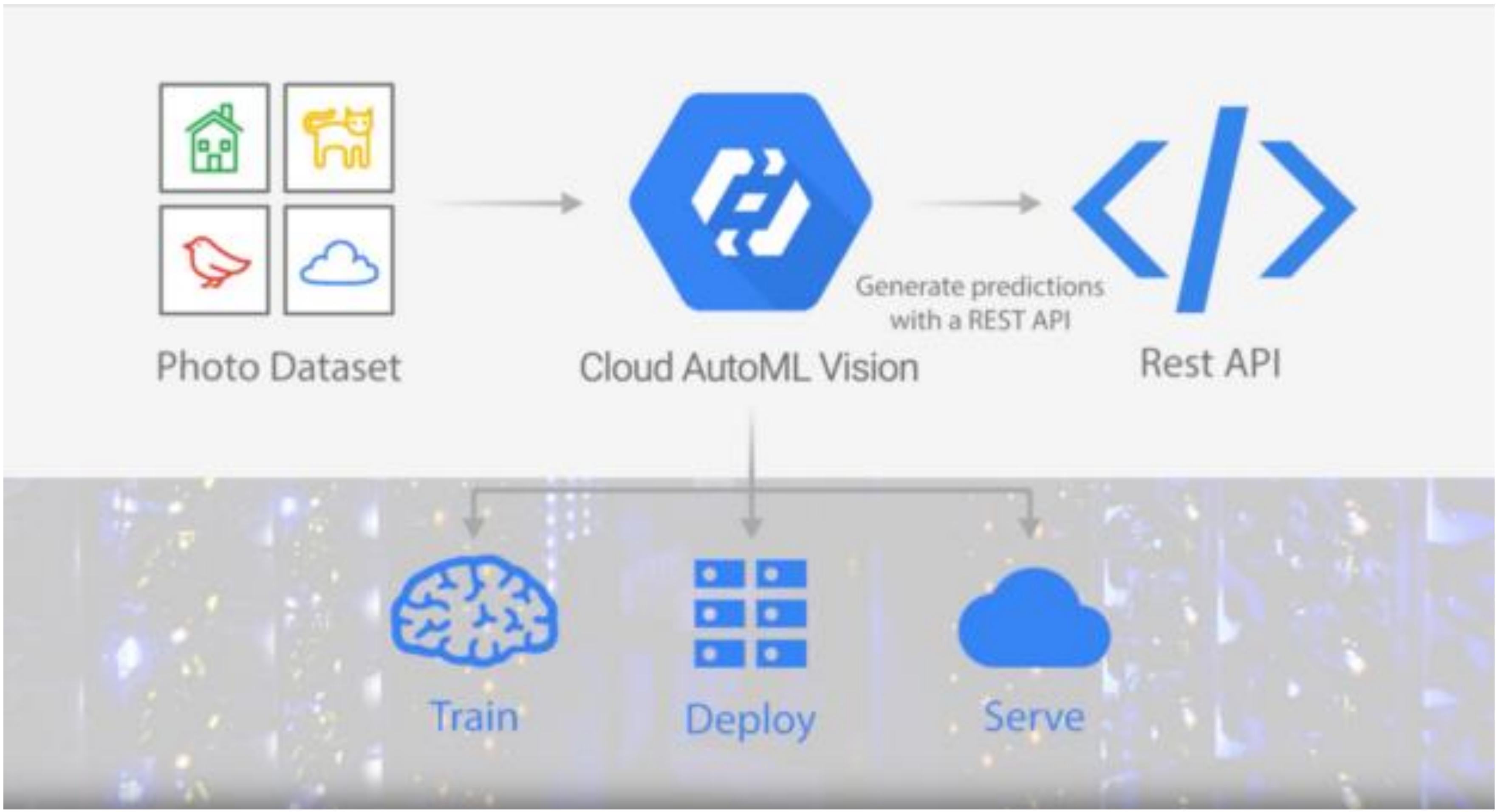


Photo Dataset
Cloud AutoML Vision
Generate predictions
with a REST API
Rest API
Train
Deploy
Serve

未來在哪裡

- 學如逆水行舟，不進則退，共勉之。