一種基於自由能與有限資源為核心的人造仿生架構之腦神經仿生計算

一、摘要

報告回顧神經計算自 1940–1950 年代以來的核心發展脈絡，從 McCulloch–Pitts 理想化神經元模型與 Hebb 法則的可塑性概念出發，經由感知機的線性限制與 XOR 瓶頸，延伸至反向傳播使多層網路可訓練化的突破，並銜接 21 世紀深度學習（CNN／RNN／LSTM／Transformer）在語音、影像與自然語言等任務上的成熟與普及。進一步地，本文說明人工神經網路作為仿生工具的合理性：其分散式處理、以權重調整承載學習、以及表徵可由資料驅動自發形成等特性，使其成為可操作、可比較的計算平台，用以在不同目標函數下產生可檢驗的內部表徵並對照神經資料。然而，標準（Artificial Neural Networks, ANN）仍存在差距，包括缺乏尖峰與膜電位等動力學、學習規則過度依賴反向傳播演算法(Backpropagation)、拓樸多固定且難以反映腦連結的多尺度結構，以及對代謝成本與全域工作空間（意識頻寬）缺乏考慮。為此，本文提出一個以**自由能與有限資源**為核心的擴充仿生機制：將**預測誤差、代謝成本、突觸與結構成本、意識工作空間頻寬、記憶容量與行為傳播**統一納入同一目標函數，並引入不規則連接圖的事前機率與慢時間尺度的拓樸可塑性，使仿生不只停留於局部結構相似，而能更完整地納入生物系統的資源與整合限制。

關鍵詞：神經計算、人工神經網路、反向傳播、深度學習、自由能原理、資訊瓶頸、工作空間、代謝成本、拓樸可塑性

二、神經計算發展的歷史概觀

（一）從神經元到邏輯單元：McCulloch–Pitts 模型與 Hebb 法則

神經計算的歷史可以追溯至 1940–1950 年代。McCulloch 與 Pitts 提出了一種理想化的神經元模型[1]：每個「經元」受多個輸入，進行加權求和，若超過某個閾值便產生輸出。形式上，這個模型可以寫成：

$$y = H\left(\sum_i \omega_i x_i - \theta\right) \tag{1}$$

其中 x_i 為輸入，ω_i 為權重，θ 為閾值，$H(\cdot)$ 為階躍函數。這個極為簡化的神經元，被證明在足夠數量與適當連接下，可以實現任意布林邏輯運算，開啟了用類神經網路進行計算的可能性。

約同一時期,Hebb 提出的**Hebbian learning**則提供了早期的學習規則雛型[2]:同時作用的神經元，其連接會被增強。在數學上，這可寫為：

$$\Delta\omega_{i;j} \propto x_i x_j \tag{2}$$

這個簡單的規則雖然與**現代深度學習的梯度下降**不同，但在概念上引入了**活動相關的可塑性**，強調突觸權重不是固定常數，而是會隨經驗而變化。

（二）感知機、限制與突破：從單層到多層網路

1950–1960 年代，Rosenblatt 提出了感知機（Perceptron）模型[3-4]，並設計了利用樣本資料調整權重的學習演算法。單層感知機可視為一個線性分類器，其決策邊界由 (3) 式所決定。然而，1969 年 Minsky 與 Papert 指出：單層感知機無法解決 XOR 等線性不可分問題，導致早期神經網路研究遇到瓶頸。

$$\sum_i \omega_i x_i + b \tag{3}$$

真正的突破出現在 1980 年代中期，Rumelhart、Hinton 等人推廣了**反向傳播（backpropagation）演算法**[5]，使多層前饋神經網路可以透過**梯度下降在誤差函數上進行優化**。對於一個多層網路，其目標是**最小化損失**，如式 (4) 所示

$$\theta \leftarrow \theta - \eta \nabla_\theta L(\theta) \tag{4}$$

其中θ表示所有權重與偏置，η為學習率。反向傳播提供了一套系統的方法來計算$\nabla_\theta L$，使得深層網路的訓練在計算上變得可行。

（三）深度學習與現代神經計算

進入 21 世紀,隨著計算資源與資料量的爆炸式成長,加上卷積神經網路(CNN)、循環網路（RNN)、長短期記憶網路（LSTM）等架構[6-8]的成熟，人工神經網路逐漸成為語音、影像、自然語言處理等領域的主力技術。後來 Transformer 架構與大型語言模型的崛起，更使大規模深度神經網路成為人工智慧的標準方案。

在這個歷程中，神經科學與工程之間的關係具有互動性：一方面，早期的網路設計與學習規則大量**借鑑了生物神經**的概念；另一方面，現代的深度學習又反過來提供了一種可訓練、可模擬的大規模網路，成為某些理論神經科學用以探索大腦計算可能形式的工具。

三、為何使用人工神經網路作為仿生工具？

（一）結構上的類比與延伸

雖然當代深度神經網路與真實神經元在細節上差距極大，但在**架構理念**上仍保留了幾個關鍵的仿生元素：
第一，資訊由多個簡單單元以**並行分布式**的方式處理，而非集中於單一計算核心；第二，計算是透過**權重的調整**來完成，類似於突觸強度的可塑性；第三，網路的表現不由單一節點所決定，而是**由整體活動模式所決定**，與大腦中群體神經元活動模式決定功能的觀點相呼應。

這種結構上的類比不要求每一個人工神經元都嚴格對應生物神經元，而是採取類仿生的立場：**保留關鍵的拓樸與計算原則，放棄微觀生物細節**，以換取可計算性與可訓練性。從這個角度看，ANN 更像是受神經啟發的統計與非線性函數逼近數學運算，但這種啟發與類比仍然提供了理解大腦計算的一個有效方式。

（二）學習與可塑性：從 Hebb 到梯度下降

使用人工神經網路的一大理由，在於它提供了一套明確的數學來描述從**經驗中學習**。在現代 ANN 中，學習通常被表達為在某個損失函數 $L(\theta)$ 進行優化：

$$\tilde{\theta} = arg\ \min_{\theta} \mathbb{E}_{(x,y)\sim\mathfrak{D}}\left[L\left(f_{\theta}(x), y\right)\right] \qquad (5)$$

這類形式雖然與真實腦內的學習機制（如 spike timing dependent plasticity, neuromodulator gating）不完全一致，但在抽象層次上反映了根據外界回饋，調整內部連接以改善行為表現這一核心概念。ANN 的優勢在於：它將這個概念變成**可實作的演算法**，可以在電腦上以大規模樣本進行反覆訓練，進而產生可觀察的行為與表徵變化，供神經科學與心理學研究做間接比較。

（三）資訊處理與表徵學習

從**資訊理論**與**統計學**的角度，人工神經網路可被視為一種**高維非線性映射**，試圖尋找從感官輸入空間到內部表徵空間的適當編碼。這些編碼往往具有**壓縮、去冗餘與特徵抽取**的性質。例如，卷積網路在影像中的早期層常會自發學出類似邊緣、方向與頻率濾波器的**權重**，與初級視覺皮層中某些神經元的感受野特性相似。雖然這種相似並不足以證明網路就是視覺皮層，但它提供了一個由**資料驅動**的表徵學習過程，顯示在適當目標函數與學習規則下，人工網路可以收斂出與生物系統部分相似的資訊處理解法。

因此，使用 ANN 作為仿生工具的理由之一，在於它提供了一個**可調整參數的實驗平台**：可以在理論上設定不同的目標（分類、預測、重建、自由能最小化等），觀察不同目標下網路自然形成的內部表徵結構，並將之與生物神經資料（如 fMRI、EEG、單細胞記錄）進行比較。這為大腦是否在做類似計算提供了一條可操作的檢驗路徑。[9]

四、限制與差距：為何標準 ANN 仍只是粗略仿生？以及可能的改進方向

儘管人工神經網路在工程應用上取得顯著成功，作為仿生模型時仍存在多層次的限制。首先，在微觀**電生理**層次，標準 ANN 中的神經元通常是簡單的非線性單元（如 ReLU、sigmoid），輸出為連續實數；真實神經元則具有**尖峰發放（spiking）、膜電位動力學、離子通道、樹突整合等複雜機制，並存在不同時間尺度的可塑性規則**。從這個角度看，ANN 與其說是神經元級的仿生，不如說是對大規模神經網路的一種平均化**粗略近似模型**。

其次，**網路拓樸與結構可塑性**的差距同樣明顯。真實大腦的連接圖**高度不規則且異質，呈現小世界結構（small-world）、多節點/核心子網路（rich-club）、腦區功能化（Functional Specialization 或 Functional Segregation）以及跨腦區的關聯湧現（Cross-regional Connectivity Emergence）等多尺度特徵**；並且這些連接會隨著發育、學習與退化而不斷重組。相較之下，多數工程上的深度網路以規則層狀架構或簡化的圖結構為主，連接關係在訓練過程中通常是權重可變但拓樸固定。即便有稀疏化或剪枝技術，多半也被視為壓縮或加速方法，而非真正的拓樸演化。這使得標準 ANN 在模擬大腦長期結構重塑時顯得力有未逮[13-14]。

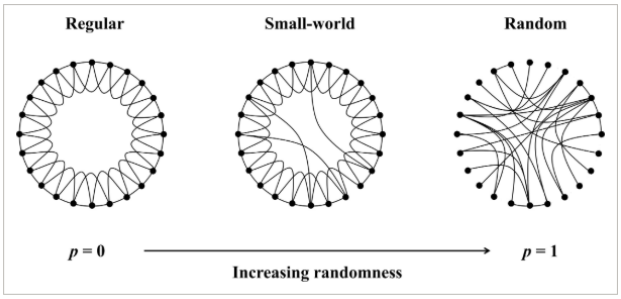

圖 1. 小世界結構與隨機性

再者，**資源限制與代謝成本**在現代 ANN 中多半以隱性或工程式的方式處理，如

限制參數數量、使用 Dropout 或權重衰減等，較少以明確的**生物能量成本**寫入目標函數。真實腦組織必須在**嚴格的代謝預算下**運作：每一個放電事件、每一個突觸的維護都需要能量，因此**少而關鍵的連接**在生物上更為合理。這提示我們，在設計仿生架構中應明確地對**活動強度**與**突觸數量**施加成本，例如在損失函數中增加對神經活動的 L2 norm 懲罰項，以及對連接權重的 L1 norm 稀疏正規，讓**節省能量與結構資源**成為網路自然追求的目標之一[15]。

最後，標準 ANN 對於**意識與有限頻寬的全域整合**往往缺乏直接刻畫。多數模型側重於輸入與輸出的映射與分層表徵，缺少一個明確的全域工作空間變數，用以承載當前**被意識到**且可在多腦區間共享的資訊。同樣地，**長期記憶**在一般深度學習中多以參數或隱狀態儲存，缺乏清楚的資訊通道觀點，無法表達在有限記憶容量下，應優先保留哪些對未來決策最關鍵的資訊。這些缺口，正是我們先前提出新架構時試圖補足之處：在傳統 ANN 之上，引入**自由能函數、有限頻寬的意識變數** y_t、有限容量的記憶變數 m_t ，以及帶有結構事前機率的不規則連接圖 G，使得仿生不再只是局部結構模仿，而是涵蓋資源與意識限制的整體計算模型[11-12]。

表 1. 標準 ANN、生物大腦與本文改進方向對照

面向	**標準 ANN（典型做法）**	**生物大腦（典型特徵）**	**本文對應的改進切入點**
單元/訊號動力學	連續值活化（ReLU/Sigmoid 等），多以靜態非線性近似神經元，通常不模型不考慮尖峰放電與時間常數	以膜電位與尖峰（spike）為主，具多時間尺度、樹突整合與雜訊特性	納入時間常數/動力學項，或採用 SNN/事件驅動模型；用時間序列目標函數約束
學習規則與信用分配	以全域誤差反傳與梯度下降為主；需要可微分、同步更新與大量標註資料	以局部可塑性（如 STDP）、調質訊號與結構性重加權為主；信用分配更偏局部與分散	引入局部學習近似（predictive coding/active inference 等）與調質因子；提升自監督/強化成分
網路拓樸與	常見為層狀前饋結構，傳播多為設計選項	大量不規則迴路、區域化與層級結構等統	以圖結構事前機率/模組化設計建模，允許

傳播迴路	（RNN/attention），拓樸多固定	計特徵	慢時間尺度的拓樸可塑性與可重構連結
資源/能量約束	資源成本多由正規化間接處理（L2/dropout 等）；能量與稀疏性通常非核心目標	代謝預算嚴格，神經活動與訊號傳遞有明確能量成本，傾向稀疏與高效表徵	在目標函數中顯式加入能量/稀疏成本（activity/weight cost），把有限資源變成可優化約束
記憶、注意力與工作空間	可用 RNN/LSTM/Transformer 實作記憶與注意力，但通常缺乏明確容量/頻寬限制的對應	工作記憶容量與注意力頻寬有限全域可用性（global workspace）常被視為意識相關瓶頸	加入意識/工作空間頻寬與記憶容量變數，並用 bottleneck/注意力限制
感知-行為閉迴路	多以監督式 loss 或外部 reward 最小化為主；行為與感知的耦合常被分離處理	以主動取樣與閉迴路控制為核心，透過行動降低不確定性與驚奇（surprise）	將行為傳播納入同一總目標（如自由能/不確定性成本），並評估探索-利用取捨

五、結論與展望：以自由能與有限資源為核心的人造仿生架構

在標準 ANN 仿生能力有限的前提下，我們可以將其視為一個起點，並在其上引入更貼近生物大腦的結構與目標。具體而言，本報告前段所討論的人工神經網路歷史與特性，可與我們設計的擴充版仿生架構結合：該架構仍以多層神經網路為基本計算單元，但在此基礎上引入一個**自由能函數**F_t，整合**解釋誤差、代謝成本、突觸與結構成本、意識頻寬、記憶容量與行為傳播與修正** 等多項考量。形式上，這個自由能可概念性地寫成：

$$F_t = E_{world} + E_{metabolic} + E_{syn} + E_{struct} + \mathcal{L}_{GW} + \mathcal{L}_{mem} - \eta \mathbb{E}[R_t] \quad (6)$$

其中E_{world}對應**網路對感官輸入的解釋誤差**（可視為預測誤差或負對數似然），$E_{metabolic}$是對**神經活動強度的懲罰**，E_{syn}與E_{struct}負責表示**突觸稀疏與不規則連接拓樸的成本**，$\mathcal{L}_{GW}$, $\mathcal{L}_{mem}$分別刻畫**有限頻寬的意識工作空間**與**有限容量的**

記憶通道，而 $\eta\mathbb{E}[R_t]$ 則代表在**當前狀態下可取得的行為回報**。網路的活動與權重更新，不再只是對單一任務損失的梯度下降，而是對這個綜合自由能的下降，較貼近生物系統在多重限制下調整自身狀態的過程，由此人工神經網路可以進一步具備幾個關鍵仿生特徵。首先，透過引入顯式的意識工作空間變數$\boldsymbol{y}_t$，並以資訊瓶頸（Information Bottleneck）的方式限制$I\left(\boldsymbol{y}_t, \boldsymbol{n}_t\right)$，網路能夠在**高維內部狀態**中選取有限數量的全域共享內容，作為決策與記憶更新的核心。其次，藉由引入記憶變數$\boldsymbol{m}_t$及資訊理論中率失真理論(Rate－Distortion)式的目標，網路可以在有限資訊通道下學習**如何在不過度耗用記憶資源的情況下，保存對未來行動最關鍵的摘要**。再者，允許連接圖 G 呈現不規則、區域模組化並可隨時間緩慢演化，搭配 L1 稀疏正規與結構上的事前機率，則使得 ANN 更接近真實大腦**亂中有序且具可塑性**的網路拓樸。這樣的架構與 Friston 的自由能原理、predictive coding、Global Workspace Theory 等理論存在自然的連結，但同時透過具體的數學設計，將**有限資源、不規則結構**與**意識與記憶**的概念落實為可優化的目標與變數。從工程角度看，它仍然是一個基於可用的現代深度學習計算概念（自動微分與梯度下降）來訓練的系統；從神經科學角度看，它則提供了一個**更接近生物限制**的計算模型，用以探索大腦如何在能量有限、連接不規則且意識頻寬受限的情況下，達成高效的資訊處理與行為調節。因此，此報告的結論可以整理如下：當前的人工神經網路在結構與學習機制上，已提供了一個有力的仿生起點，但若要更貼近真實大腦，必須進一步將資源限制、不規則拓樸與意識記憶機制納入同一個自由能框架之中。這樣的擴充模型並非現有神經科學的既定主流，而是一種建構性提案：試圖在工程可行性與生物合理性之間取得新的平衡，使 ANN 不僅是模式識別器，也逐步成為具有解釋力的仿生神經系統。未來，透過漸進式的簡化模型與模擬實驗，我們有機會驗證此類架構在實際行為與神經資料上的表現，並藉此反向推進我們對大腦與意識本質的理解以期形成良性發展循環[10, 16]。