

三、研究計畫內容（以中文或英文撰寫）：

（一）研究計畫之背景。請詳述本研究計畫所要探討或解決的問題、研究原創性、重要性、預期影響性及國內外有關本計畫之研究情況、重要參考文獻之評述等。如為連續性計畫應說明上年度研究進度。

這幾年的深度學習的許多成功案例在產官學研引起極大的震撼，發展機器學習和人工智慧，不僅是各大 IT 業者的首要任務，也被許多國家列為未來經濟發展及科技推動的重點政策。我國除了可以投入在深度學習的開發上，其實，我們可以在更廣義的機器學習上有所突破。

從 1950 年代，各界就期望能夠設計出一部像人類大腦一樣，能夠學習及具有智慧的機器，如此一來，許多複雜難解或有生命危險等高難度的工作，便可以交由此等智慧型的機器來完成。因此，促成了機器學習和人工智慧的各自發展。這段期間，歷經了多次的瓶頸突破，首先是在 1950 年代的感知器的問世，揭開類神經網路研究的序幕，但由於它的線性化的構造無法解決較複雜的非線性問題，導致類神經網路進入停滯期。直至 80 年代的倒傳遞演算法 (backpropagation algorithm) 的出現，解決了多層感知機訓練的問題，類神經網路再度又開始被大量地應用於許多不同的領域。由於受限於當時的計算資源和特徵選取的問題，這些應用大都採用淺層的類神經網路架構，雖然已能成功解決許多問題，但一些複雜的電腦視覺及行為解讀的能力仍力有未逮。

近幾年來，由 Geoffrey Hinton、Yann Lecun 和 Yoshua Bengio 等人所帶動的深度學習的許多成功應用，吸引了各個領域的目光，特別是 AlphaGo 打敗人類圍棋頂尖棋手的新聞，讓一般大眾都為之著迷及震撼等，再度將機器學習和人工智慧的研究推上另一番高峰。其實，深度學習仍算是一種多層的類神經網路，只是傳統多層的類神經網路遭遇到兩個主要問題：(1) 訓練演算法的有效性：若全用倒傳遞演算法 (backpropagation algorithm) 來訓練多層的類神經網路，通常訓練效果並非十分理想；(2) 特徵的有效性：傳統方法解決各式的問題時，都需要該應用領域的專家知識來決定何種輸入特徵較適合來解該類的問題（如：語音辨識時就用 mel-cesptrum 係數，影像處理時就常用 Wavelet Transform、Gabor Transform 等），然後，再搭配一個合適的分類器 (classifier) 予以分類。但這些專家選取的特徵並非總是最合適或最有效的特徵，導致訓練結果有時只能差強人意或有侷限性。直到 2006 年左右，Hinton 等人提出各種開創性的訓練方式及架構（如：Deep Brief Networks (DBN), Convolutional Neural Network (CNN)等）[1]-[6]，才讓多層的類神經網路的有效性被充分地發揮，也因此，深度學習開始吸引各個領域及世界各知名大廠的注意力，蓬勃發展至今。目前，有許多 Deep Learning Toolbox 提供大家使用[7]，其中廣被採用的有 TensorFlow、Theano、Caffe、DeepLearnToolbox、CudaConvnet、Deeplearning4J、MSHadow 和 Gnumpy 等。

深度學習成功的關鍵在於：(1) 網路的特殊架構、(2) 巨量的訓練資料、(3) 進展神速的硬體計算資源和 (4) 許多開源工具的釋出。由於產官學研的相繼投入，導致這方面的研究蓬勃發展，相對地，要能與各大 IT 廠商有所競爭及差異化變得日益困難。台灣在缺乏大數據的環境之下，要在深度學習或人工智慧的發展趨勢之下能夠迎頭趕上甚至超越，一方面是可以考慮在硬體製造的基礎上，發展低耗能及為深度學習量身打造的高速計算晶片；另一方面，就是要擴大深度學習的應用範疇和提出新的機器學習演算法，以避開從後苦苦追趕大廠的窘境。這樣的發展策略勢必就要選擇合適且有潛力的應用範疇，然後，針對所聚焦的應用範疇來開發新的網路架構和學習演算法。

此計畫是一個三年期的計畫，整體的總目標是要開發一個新的機器學習演算法，並將它可應用於不同情境之姿態辨識系統中，其中一個應用情境是獨居老人的居家生活動作辨識，另一個應用情境是孩童的學習姿態的辨識。因為老人和孩童是家庭成員中最須關注的對象，有其必要性和未來發展性。

類神經網路(neural networks)或學習機器 (learning machines) 的性能很大程度上取決於它是否能

夠成功地從訓練資料集中學習相應的映射。在類神經網路應用於解決各種問題上，有兩種學習問題亟需解決：1) 參數學習 (parameter learning) 和 2) 結構學習 (structure learning) [8]。參數學習涉及到類神經網路中連接權值的更新，而結構學習重點在於確定網路結構，如：層數和每一層神經節點。這兩種學習問題既可以同時解決，也可以分開解決。各種方法都有其自身的吸引力和局限性。大多數類神經網路（例如，多層感知器和 RBF 網路）從小的網路架構開始，然後使用離線更新規則來更新網路的連接權重以實現某種性能評估標準(如：最小均方誤差)。如果性能標準不能達到預先規定的成效，那麼網路架構將會再增加，網路將被重新初始化和重新訓練。由於加權初始化、網路規模、學習規則和數據映射的複雜性等問題總是混雜在一起，所以網路架構的增加並不意味著網路架構較大的網路的性能絕對會超越架構較小的網路尺寸。另一種方法是首先使用結構學習算法來找到合適的網路規模，然後使用參數學習規則來微調連接權值。模糊自適應學習控制網路 (FALCON) 是採用上述方法的典型神經模糊系統之一 [9]。由於 FALCON 在訓練數據在線生成的情況下不能很好地運作，為了賦予網路在線學習能力，提出了將參數學習的反向傳播算法和模糊 ART 結合學習的 FALCON-ART [10]。在一些應用中，系統應該能夠學習新的信息而不會破壞以前獲得的知識，這個能力是所謂的「增量學習能力 (incremental learning)」。可以學習新的信息而不忘記以前獲得的知識，然而，這樣的能力會遭遇所謂的「穩定性 - 可塑性的困境 (stability-plasticity dilemma)」[11]。具有增量學習能力的類神經網路為解決困境提供了可能的解決方案。模糊 ARTMAP 系統 [12]，模糊最小最大類神經網路 [13] 和 HS-ARTMAP 網路 [14] 是具有增量學習能力的三種類神經網路。

最近，Huang 等人提出了一種「極限學習機器 (extreme learning machine, 簡稱 ELM)」來快速生成一個前饋神經網路的單隱層 [15]-[20]。ELM 為解決參數學習和結構學習提供了一個有吸引力的解決方案。在 ELM 中，隨機選擇隱藏節點的數量和活化函數 (activation function) 的權重，但是通過使用 Moore-Penrose 廣義逆算子分析確定輸出權重 (將隱藏節點鏈接到輸出節點)。ELM 可以顯著地減少訓練單隱層前饋神經網路所需的時間；然而，與傳統的學習算法 [21]-[22] 相比，由於隨機確定輸入權重和隱藏偏差，ELM 可能需要更多的隱藏節點。在去年我們提出了「以 Jacobian Matrix 為基礎之學習機器 (Jacobian Matrix-based learning machine, 簡稱 JMLM)」[23]，JMLM 提出了另一個簡單而有效的解決方案來解決參數學習和結構學習的問題。提出的解決方案背後的基本想法是使用泰勒級數展開來從一組訓練資料中建立該有的輸入-輸出映射函數。眾所周知，如果一個輸入變量非常接近一個特定的採樣點，那麼它的相應的函數值可以通過在該特定點上的一階泰勒級數展開來近似。泰勒級數展開式的概念可以推廣到向量值函數，而此時的一階導數是被命名為 Jacobian Matrix (雅可比矩陣)，所提出的 JMLM 的性能完全取決於採樣點的選擇以及所選採樣點處的 Jacobian Matrix 的成功計算。

由於原先的 JMLM 屬於淺層的網路架構，雖然可以得到很高的辨識率，但付出的代價是有時會需要產生很多的節點，增加了複雜度。而另一方面，由於深層的網路架構可以呈現複雜的階層式特徵的功能，但需要足夠的硬體資源和好的訓練演算法才能發揮功能。因此，此計畫的第一年是要結合 JMLM 和深度學習的多層架構的優點來開發「以 Jacobian Matrix 為基礎之深度學習機器，簡稱 Deep JMLM」，希望此 Deep JMLM 可以逐步搜索出一組適當的採樣點及層數，使得它們的相對應的函數值可以很快地透過 Jacobian Matrix 的計算得到，並因此而達到降低均方誤差 (mean squared error) 的訓練目標。若此 Deep JMLM 開發成功，緊接著就會將它應用於不同情境之姿態辨識系統中。

未來，世界各地都陸續要面臨高齡化所帶來的衝擊，針對高齡化社會的居家照護會是一個值得開發研究的議題，且這方面的成果除了可以解決社會問題之外，也具產業效益。我國在今年時，老年人口比例已經正式超越百分之十四，也已邁入高齡社會 [24]。現今社會，多數子女都出外工作或自主家庭，因此，獨居老人的照護逐漸地就顯得相當需要被重視。有時子女只能一個星期或者是更久的時間才能回家關心或照護父母，雖然，現今視訊軟體非常方便，每天都可連繫，但父母有時不願意讓子女擔心身體狀況，所以，都會盡量隱藏身體狀況。因此，如果有個系統能夠對獨居老人的

日常生活活動量做客觀的評估，又能提供跌倒偵測的功能，那將會是一個有實際需求之照護系統。此外，此系統也能提供保險公司或長照系統更有效地進行資源調配，對台灣的照護體系將有不小的幫助。目前常見的日常生活功能評估量表有巴氏量表、祈氏量表以及柯氏量表，其中較常用的為巴氏量表，此量表常被做為進行居家護理、申請外籍看護、申請長照服務及觀察臨床變化時的評估依據。日常生活功能評估量表因其執行簡單和成本低廉故被廣泛地使用在臨床上，不過需透過醫護人員目視評估填寫，其缺點是無法實時地量測得知病人的狀況。所以，本計畫聚焦於將機器學習應用於如何有效且即時地對獨居老人的活動量進行客觀的量測研究，在此第二年的計畫中要針對獨居老人開發一套居家生活動作辨識系統，預計要能辨識十種不同之居家生活動作（如：站立、彎腰、蹲姿、跌倒等），最後此系統可提供一個可供分析的活動量評估報告。

過去幾年來，已有許多有關動作辨識或姿態辨識的研究，各有其優缺點。目前的動作辨識主要利用人體剪影（human silhouettes）、光流法（optical flow）[25]、影像梯度（image gradients）以及區域時空特徵 [26]等，利用這些特徵在以機器學習進行分類，都能得到還算不錯的效果。在動作辨識時主要可以分成兩個階段，第一個階段為特徵截取，此階段將從攝影機所得之影像以不同的特徵表示方式表示出來用以代表該動作；第二個階段為將先前各個動作的特徵表示以分類器進行訓練，而在特徵截取的方面有兩種不同的特徵截取辦法，分別為全局特徵表示（global representations）以及區域特徵表示（local representation），各代表不同的分析策略。全局特徵表示的方式為由上至下的方式（top down approach），而區域特徵表示則為由下至上的方式（bottom up approach）。全局特徵表示方式通常是透過背景相剪法（background subtraction）的來得到目標區塊（interest of region），在經過人體剪影（human silhouettes）和光流法（optical flow）的方式來得到全局特徵，但相剪時易受到雜訊和不同的視角而受到影響，常見的全局特徵動作辨識方式有以下兩種：Bobick 和 Davis 利用人體剪影進行動作辨識 [27]，論文中先產生 motion-energy image（MEI）和 motion-history image（MHI），此兩種可視為不同的時空模板（template），然後，需要識別的動作在與存在於資料庫中已知動作的兩種模板做比對即可完成分類。Danafar 和 Gheissari 以光流法來進行動作辨識 [28]，此篇論文先利用框出有人的範圍，利用比例將人分成頭、軀幹和腳部，再以垂直和水平的光流法來進行處理，以此可以得到不同的角度的光流場的特徵，再利用 Support Vector Machine（SVM）進行分類。

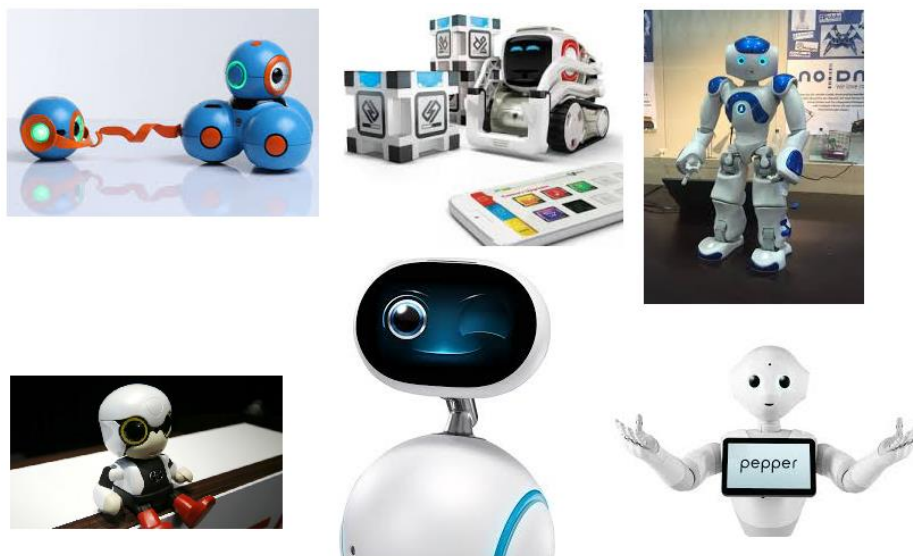
近年來，姿態辨識系統常採用骨架資訊為基礎。一種普遍的骨架偵測方式是透過 Kinect 的深度資訊及 SDK，但由於 Kinect 的深度資訊的有效範圍在四公尺之內，不適合在居家的客廳環境應用。近年來，因為深度學習的發展迅速，使得從 2 維影像可以實時地估測出骨架。在 2016 年的 MSCOCO Keypoints Challenge 由 Zhe Cao 等人所提出以 Part Affinity Fields（PAF）由下至上的方法（bottom up approach）[29]，方法是以卷積類神經網路學習 Part Confidence Maps，將影像中的每個關節點以及其相連的肢體連結標記出，再利用 Part Affinity Fields 做肢體連接即可完成姿態骨架偵測（如圖一所示），其演算法可以實時的做姿態估測並且在人數增加的情況下依然能夠保持穩定的 FPS，已具備高度的實用性。我們第二年的計畫會基於此研究成果向上發展出一套居家生活動作辨識系統。



圖一、姿態骨架偵測 [29]

家庭成員中另一個最需被關注的就是孩童了，當父母親不能陪伴在小孩身邊時，父母親都關心小孩在自己的房間念書時，是否真的有專心在念書？還是東摸西摸地打發時間而已？抑或是趴在書桌上睡著了？若是小孩的確有認真念書，雖然會感到很欣慰，但也會擔心小孩是否有保持適當的閱讀距離？閱讀姿勢正確與否？許多的父母親真的關心小孩的**專注力**和**視力保健問題**，雖然，市面上已有許多陪伴型機器人的產品問世（如圖二所示），但都沒有具備上述功能。所以，此計畫的第三年是要開發孩童的學習姿態辨識系統來陪伴孩童讀書，此學習姿態辨識系統具備**專注力偵測**和**閱讀姿態偵測**兩項重要功能，其目標是要減輕父母親的負擔和掌握孩童學習的狀態。

- (1) **專注力偵測功能**：小孩能專注在念書或寫作業時，父母親是會感到很高興，但也希望能每隔一段專注時間後，要小孩記得起身走動走動。若是長時間都沒在念書或睡著了，會希望提醒小孩該專心了。
- (2) **閱讀姿態偵測功能**：每當孩童在書桌上沉浸於書本或是使用電腦時，經常會不自覺地與書本或是螢幕越靠越近，這時近視的可能性便大大地提升了，更何況孩童往往不具備保護眼睛的觀念，而大人又無法時時刻刻待在小孩旁邊。一旦孩童眼睛與書本或螢幕靠得太近，此系統功能會馬上發出警示，如此一來，便降低了因為閱讀姿勢不良所造成的近視機率，孩童健康成長，家長也放心。此外，有些長時間累積而來的壞習慣，例如駝背、手撐頭等不良坐姿，都很有可能使孩童未來成長發生問題，因此本功能包含了坐姿監測，讓孩童從小開始培養好習慣。



圖二、各類機器人商品

參考文獻

- [1] G. E. Hinton, S. Osindero, and Y. Teh, ,“A fast learning algorithm for deep belief nets,” Neural Computation, vol. 18, pp. 1527-1554, 2006.
- [2] Y. Bengio, P. Lamblin, D. Popovici and H. Larochelle, “Greedy Layer-Wise Training of Deep Networks,” in J. Platt et al. (Eds), Advances in Neural Information Processing Systems, vo. 19, pp. 153-160, MIT Press, 2007.
- [3] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, “Efficient Learning of Sparse Representations with an Energy-Based Model,” in J. Platt et al. (Eds), Advances in Neural Information Processing Systems, MIT Press, 2007.
- [4] G .E. Hinton, "Deep belief networks". Scholarpedia, vol. 4, no. 5, pp. 5947, 2009.
- [5] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural Networks, vol. 61, pp.85-117,

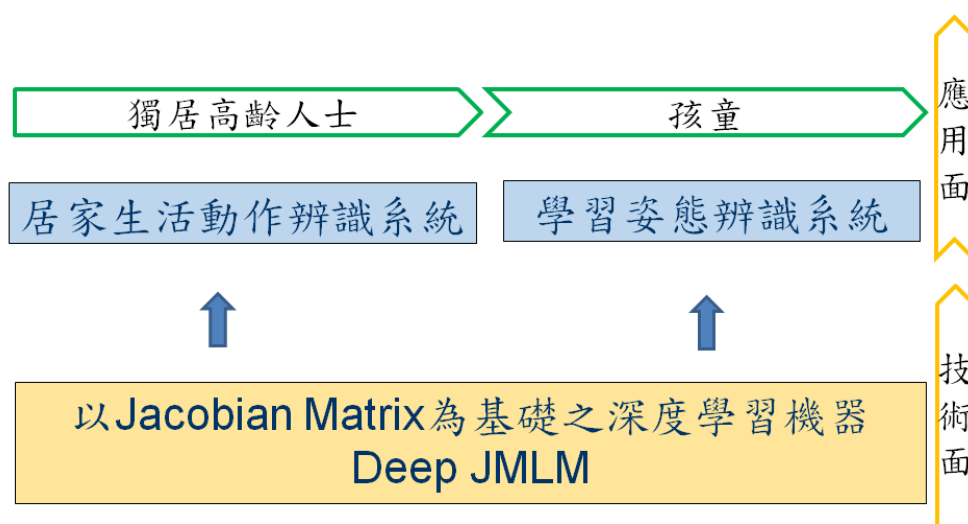
2015.

- [6] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition.," Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.
- [7] Deep Learning Software links, [Online] Available: http://deeplearning.net/software_links/ .
- [8] C. -T. Lin and C. S. G. Lee, Neural Fuzzy Systems: A Neuro-fuzzy Synergism to Intelligent Systems, Prentice-Hall International, Inc., 1996.
- [9] C. -T. Lin and C. S. G. Lee, "Neural-network-based fuzzy logic control and decision system," in IEEE Transactions on Computers, vol. 40, no. 12, pp. 1320-1336, 1991.
- [10] C. -J. Lin and C. -T. Lin, "An ART-based fuzzy adaptive learning control network," Proceedings of IEEE International Conference on Fuzzy Systems, Vol. I, pp. 1-6, 1994.
- [11] S. Grossberg, "Nonlinear neural networks: principles, mechanism, and architectures," Neural Networks, vol. 1, no. 1, pp. 17-61, 1988.
- [12] G. A. Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, and D. B. Rosen, "Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps," IEEE Trans. on Neural Networks, vol. 3, pp. 698-713, 1992.
- [13] P. Simpson, "Fuzzy min-max neural networks— Part 1: Classification," IEEE Trans. on Neural Networks, vol. 3, no. 5, pp.776-786, Sept. 1992.
- [14] M. C. Su, J. Lee, and K. L. Hsieh, "A new ARTMAP-based neural network for incremental learning," Neurocomputing, vol. 69, pp. 2284-2300, 2006.
- [15] G. -B. Huang, Q. -Y. Zhu, and C. -K. Siew, "Extreme learning machine: A new learning scheme of feedforward neural networks," in Proc. IJCNN, Budapest, Hungary, vol. 2, pp. 985–990, 2004.
- [16] G. -B. Huang, Q. -Y. Zhu, and C. -K. Siew, "Extreme learning machine: Theory and applications," Neurocomputing, vol. 70, no. 1–3, pp. 489–501, 2006.
- [17] G. -B. Huang, L. Chen, and C.-K. Siew, "Universal approximation using incremental constructive feedforward networks with random hidden nodes," IEEE Trans. on Neural Networks, vol. 17, no. 4, pp. 879–892, 2006.
- [18] G. -B. Huang, S. Song, J. N. D. Gupta, and C. Wu, "Semi-supervised and Unsupervised Extreme Learning Machines," IEEE Transactions on Cybernetics, vol. 44, no. 12, pp. 2405-2417, 2014.
- [19] G. -B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme Learning Machine for Regression and Multiclass Classification," IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics, vol. 42, no. 2, pp. 513-529, 2012.
- [20] Z. Bai, G. -B. Huang, D. Wang, H. Wang, and M. B. Westover, "Sparse Extreme Learning Machine for Classification," IEEE Transactions on Cybernetics, vol. 44, no. 10, pp. 1858-1870, 2014.
- [21] R. Rajesh and S. Prakash, "Extreme learning machines- a review and state-of-the-art," International Journal of Wisdom based Computing, vol. 1, no. 1, pp. 35-49, 2011.
- [22] Q. -Y. Zhu, A. K. Qin, P. N. Suganthan, G. -B. Huang, "Evolutionary Extreme Learning Machine," Pattern Recognition, vol. 38, pp. 1759-1763, 2005.
- [23] M. C. Su, Y. Z. Hsieh, C. -H. Wang, and P. -C. Wang, "A Jacobian Matrix-based Learning Machine and Its Applications in Medical Diagnosis," *IEEE Access*, pp. 1, 2017 (SCI).
- [24] 06 年 第 10 週 內 政 統 計 通 報 . [Online]. Available: http://www.moi.gov.tw/stat/news_content.aspx?sn=11735. [Accessed: 4-Jul-2017].
- [25] B. D. Lukas, and T. Kanade, "An iterative image registration technique with an application to stereo

- vision,” *7th International Joint Conference on Artificial Intelligence*, pp. 674-679, 1981.
- [26] I. Laptev, and T. Lindeberg, “Local Descriptors for Spatio-Temporal Recognition,” *Spatial Coherence for Visual Motion Analysis*, Vol. 3667, pp. 91-103, 2004.
- [27] A. F. Bobick, and J. W. Davis, “The recognition of human movement using temporal templates,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23 No. 3 pp. 257-267, 2001.
- [28] S. Danafar, and N. Gheissari, “Action recognition for surveillance applications using optic flow and SVM,” *Proceedings of the Asian Conference on Computer Vision*, Vol. 4844, pp. 457-466, 2007.
- [29] Z. Cao, T. Simon, S. E. Wei, and Y. Sheikh, “*Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields*,” *Conference on Computer Vision and Pattern Recognition*, 2017.

(二) 研究方法、進行步驟及執行進度。請分年列述：1.本計畫採用之研究方法與原因及其創新性。2. 預計可能遭遇之困難及解決途徑。3.重要儀器之配合使用情形。4.如為須赴國外或大陸地區研究，請詳述其必要性以及預期效益等。

此三年期的計畫 (如圖三所示)，預計第一 year 要開發「以 Jacobian Matrix 為基礎之深度學習機器，簡稱 Deep JMLM」，然後，基於此 Deep JMLM 來開發可適用於兩種家庭成員的姿態辨識系統。第二 year 要針對獨居老人開發一套居家生活動作辨識系統，預計要能辨識十種不同之居家生活動作，最後提供一個可供分析的活動量評估報告。而第三年的計畫目標對象則是針對學童開發一套學習姿態辨識系統，可以全程掌握孩童念書或上課時的專注的程度，並能適時提醒孩童眼睛是否距離書本太近或是有不良之閱讀坐姿(如：駝背、手撐頭等)，其目標是要減輕父母親的負擔和掌握孩童學習的狀態。



圖三、計畫架構

第一年：「以 Jacobian Matrix 為基礎之深度學習機器，簡稱 Deep JMLM」

類神經網路或學習機器的訓練精神就在於如何從收集到的訓練資料， $(\mathbf{x}_1, \mathbf{F}(\mathbf{x}_1))$, ...,

$(\mathbf{x}_N, \mathbf{F}(\mathbf{x}_N))$ ，透過參數的調整來找到隱藏在其中的輸入/輸出的映射函數，此訓練得到的映射函數可被用來處理圖樣識別或函數逼近的問題。

我們這個第一年的計畫是要提出一個整合了 JMLM 和深度學習的優點的 Deep JMLM。整個 Deep JMLM 是根植於多維的泰勒展開式數學定理，所以，我們很快地回顧泰勒展開式。根據泰勒展開式數學定理，一個純量函數 $f(x)$ 可以用一個無限的和項來表示，這個和可以從函數在一個點 p 的導數值上計算出來，這個值接近於 x ，如下：

$$f(x) = f(p) + \frac{f'(p)}{1!}(x-p) + \frac{f''(p)}{2!}(x-p)^2 + \frac{f^{(3)}(p)}{3!}(x-p)^3 + \dots \quad (1)$$

如果點 p 非常接近 x ，則高階導數可以被省略，因此，函數 $f(x)$ 可以用以下線性方式近似：

$$f(x) \cong f(p) + f'(p)(x-p) \quad (2)$$

線性泰勒多項式可以推廣到多維度向量值函數，其推廣情況如下：

$$\mathbf{F}(\mathbf{x}) \cong \mathbf{F}(\mathbf{p}) + \mathbf{J}(\mathbf{p})(\mathbf{x}-\mathbf{p}) \quad (3)$$

其中 x 和 p 是 n 維向量， $\mathbf{F}(\mathbf{x}): \mathbb{R}^n \rightarrow \mathbb{R}^m$ 是多維度向量值函數， \mathbf{J} 是 Jacobian matrix（雅克比矩陣）。雅可比矩陣 \mathbf{J} 是向量值函數 $\mathbf{F}(\mathbf{x})$ 的所有一階偏導數所構成的矩陣，如下所示：

$$\mathbf{J} = \begin{bmatrix} \frac{\partial F_1}{\partial x_1} & \dots & \frac{\partial F_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial F_m}{\partial x_1} & \dots & \frac{\partial F_m}{\partial x_n} \end{bmatrix} \quad (4)$$

接下來的問題是要如何解決雅可比矩陣在某一點取樣點的估計值。假設我們有 $N+1$ 個數據對， $(\mathbf{x}_1, \mathbf{F}(\mathbf{x}_1)), \dots, (\mathbf{x}_N, \mathbf{F}(\mathbf{x}_N)), (\mathbf{p}, \mathbf{F}(\mathbf{p}))$ 。有以下兩種方法可從 $N+1$ 個數據對來估計雅可比矩陣：

方法一：最小均方差演算法 (The Least-Mean-Square (LMS) Algorithm)

採法採用迭代方式來估算雅可比矩陣，其公式如下：

$$\mathbf{J}(\text{new}) = \mathbf{J}(\text{old}) - \eta \frac{\partial E}{\partial \mathbf{J}} = \mathbf{J}(\text{old}) - \eta (\Delta \mathbf{F} - \mathbf{J}(\text{old})\Delta \mathbf{p})(\Delta \mathbf{p})^T \quad (5)$$

其中

$$\Delta \mathbf{F} = \mathbf{J}(\mathbf{p})\Delta \mathbf{p} = \begin{bmatrix} J_{11} & \dots & J_{1n} \\ \vdots & \ddots & \vdots \\ J_{m1} & \dots & J_{mn} \end{bmatrix} \Delta \mathbf{p} \quad (6)$$

$$\Delta \mathbf{p} = \begin{bmatrix} x_1 - p_1 \\ \vdots \\ x_n - p_n \end{bmatrix} \text{ and } \Delta \mathbf{F} = \begin{bmatrix} F_1(\mathbf{x}) - F_1(\mathbf{p}) \\ \vdots \\ F_m(\mathbf{x}) - F_m(\mathbf{p}) \end{bmatrix} \quad (7)$$

方法二：Moore-Penrose 廣義反矩陣法 (The Moore-Penrose generalized Inverse)

$$\mathbf{J}_{m \times n} = \Delta \mathbf{F}_{m \times N} (\Delta \mathbf{p}_{n \times N})^T ((\Delta \mathbf{p}_{n \times N} (\Delta \mathbf{p}_{n \times N})^T)^{-1})^T \quad (8)$$

其中

$$\begin{bmatrix} \Delta F_1^1 & \cdots & \Delta F_1^N \\ \vdots & & \vdots \\ \Delta F_m^1 & \cdots & \Delta F_m^N \end{bmatrix}_{m \times N} = \begin{bmatrix} J_{11} & \cdots & J_{1n} \\ \vdots & \ddots & \vdots \\ J_{m1} & \cdots & J_{mn} \end{bmatrix}_{m \times n} \begin{bmatrix} \Delta p_1^1 & \cdots & \Delta p_1^N \\ \vdots & & \vdots \\ \Delta p_n^1 & \cdots & \Delta p_n^N \end{bmatrix}_{n \times N} \quad (9)$$

$$\Delta \mathbf{F}_{m \times N} = \mathbf{J}_{m \times n} \Delta \mathbf{p}_{n \times N} \quad (10)$$

有了上述如何從 $N+1$ 個數據對， $(\mathbf{x}_1, F(\mathbf{x}_1)), \dots, (\mathbf{x}_N, F(\mathbf{x}_N)), (\mathbf{p}, F(\mathbf{p}))$ ，估計出雅可比矩陣之解決方案後，我們預計會提出以下的學習演算法來建構一個可用於圖樣識別或函數逼近的 Deep JMLM。整體訓練算法構想是從訓練資料的加以分群開始，然後遞增地將新的採樣點添加到學習機器中以改善其性能。Deep JMLM 的訓練算法的構想如下：

步驟 1：初始化 - 指定最小性能標準 θ 和最大採樣點數量 K_{max} 。初始化索引 K （即訓練數據集中當前採樣點或群聚的數量）為 1。

步驟 2：分群 - 使用 K 均值算法將輸入數據集 $X = \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$ 分成 K 群聚， $C_i = \{\mathbf{x}_1^i, \mathbf{x}_2^i, \dots, \mathbf{x}_{N_i}^i\}$ for $i = 1, \dots, K$ 其中 $X = C_1 \cup C_2 \cup \dots \cup C_K$ 和 $N = \sum_{i=1}^K N_i$ 。對於群聚 C_i ，其群聚中心要用與其最接近的輸入數據替換，並且我們使用 \mathbf{p}_i 來表示第 i 個被置換群聚的中心。將輸入數據的期望輸出分配為替換的聚類中心的期望輸出，並表示為 \mathbf{d}_i^p 。這些 K 被取代的群聚中心可以作為計算雅可比矩陣的取樣點。

步驟 3：雅可比矩陣的計算 - 對於第 i 個群聚 C_i ， \mathbf{p}_i 及其對應的期望輸出 (i.e., $(\mathbf{x}_1^i, \mathbf{d}_1^i), \dots, (\mathbf{x}_{N_i}^i, \mathbf{d}_{N_i}^i)$, and $(\mathbf{p}_i, \mathbf{d}_i^p)$)，我們可以使用 LMS 算法，如公式 (5) 式或 Moore-Penrose 廣義反矩陣法 (8) 計算其對應的雅可比矩陣 \mathbf{J}_i 。根據許多模擬結果，Moore-Penrose 廣義反矩陣法是較被推薦的，因為它可以比 LMS 算法更快，更有效地計算 Jacobian 矩陣。

步驟 4：輸出計算 - 對於輸入數據 \mathbf{x}_i ，我們尋找離被替換的群聚中心 (e.g., \mathbf{p}_{i^*}) 最近的輸入數據 \mathbf{x}_i 。這個具有 K 個採樣點的訓練後的 Deep JMLM 的預測輸出的計算方式如下：

$$\mathbf{F}(\mathbf{x}_i) = \mathbf{F}(\mathbf{p}_{i^*}) + \mathbf{J}_{i^*}(\mathbf{p}_{i^*})(\mathbf{x}_i - \mathbf{p}_{i^*}) = \mathbf{p}_{i^*}^d + \mathbf{J}_{i^*}(\mathbf{p}_{i^*})(\mathbf{x}_i - \mathbf{p}_{i^*}) \quad (16)$$

函數 $\mathbf{F}(\mathbf{x}_i)$ 給出了從數據點 \mathbf{x}_i 到由 $\mathbf{F}(\mathbf{x}_i) = 0$ 定義的超平面的距離的代數測度。如果數據點 \mathbf{x}_i 位於超平面的正側（即，其相應的法向量的超平面的相同側），則函數 $\mathbf{F}(\mathbf{x}_i)$ 的值是正的，反之亦然。如果訓練問題是一個函數逼近問題，則採用平均平方誤差來驗證訓練的 JMLM 的性能：

$$E_{av} = \frac{1}{N} \sum_{i=1}^N \|d_i - F(x_i)\|^2 \quad (17)$$

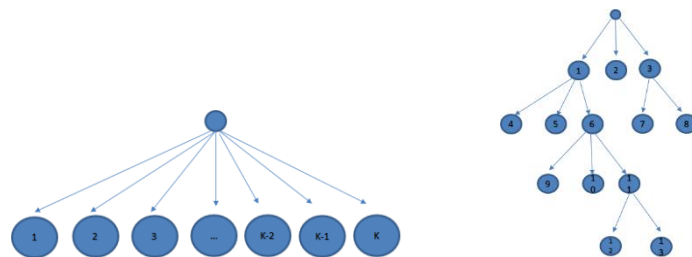
如果訓練問題是模式識別問題，則可以採用平均方差或整體識別率來驗證訓練的 JMLM 的性能。假設學習問題是解決 P 個類別的分類問題。我們需要將 $F_i(x_i)$ 的輸出值離散成 P 個區間，並且如果 $F_i(x_i)$ 落在對應於該類的區間內，則將類型 x_i 分配給類。

步驟 5：終止 - 如果平均平方誤差小於預先設定的閾值 θ 或採樣點數量達到其最大值 K_{max} ，則終止訓練過程。否則， $K = K + 1$ 並轉到步驟 2。

預計可能之困難與解決途徑：

1. 困難：何時增加下一層節點的決定？

解決途徑：原先的 JMLM 是採用逐漸地增加群聚數目的方式來降低平均平方誤差小於預先設定的閾值 θ ，這樣的做法會導致形成如圖四所示的淺層(只有一層)樹狀結構的學習機器，這樣會導致將輸入空間被切割得太細瑣，可能會產生推廣能力 (generalization capability) 的下降。所以，我們預計會採用在每一個節點限制可連接節點數目的方式來決定何時要增加下一層的節點，依此原則，不斷地增加層數直致均方誤差降到預定的設定。如此建立的 JMLM 就會具備多層的架構，希望一方面能降低均方誤差，另一方面可以克服推廣能力下降的缺點。



圖四、JMLM 與 Deep JMLM 的差異比較

2. 困難：如何在廣度和深度的折衷上做好的取捨？

解決途徑：只有一層的 JMLM 的優點是可以從訓練後的 JMLM 的參數中萃取出如下的解釋規則：

If x is close to $(0.12, 5.41, 0.26, 0.19, 0.24, 9.79, 8.46, 7.84, 2.54)^T$
Then $F_1(x) = 1 - 0.42(X_1 - 0.12) - 0.022(X_2 - 5.41) - 0.32(X_3 - 0.26)$
 $- 0.032(X_4 - 0.19) + 0.23(X_5 - 0.24) - 0.031(X_6 - 9.79) - 0.0078(X_7 - 8.46)$
 $+ 0.032(X_8 - 7.84) - 0.0055(X_9 - 2.54)$

根據上述判斷規則，使用者可以很快地得知那些特徵在決策時比較重要？但只有一層的代價就是會將輸入空間切割得太細瑣，導致規則太多。而此次要研發的 Deep JMLM 改善了切割得太細瑣的缺點，但付出的代價是產生巢狀 (nested) 的規則，這樣的巢狀規則會較不利直覺的解讀。所以，必須在廣度和深度上的做折衷才能有好的推廣利和可解釋性。初步，我們可能會借助閾值 θ 和節點數的上限來做控制。

第二年：獨居老人所開發的居家生活動作辨識系統

近年來拜醫療科技的進步，台灣社會面臨人口結構高齡化的問題，而隨這年輕子女的外移，獨居老人照護問題也比以往更加需要被關注，如何有效且即時的對獨居老人的活動量進行量測是目前很重要的一個議題。今年的計畫目標是以所開發完成之 Deep JMLM 為基礎，研發一套居家生活動作辨識

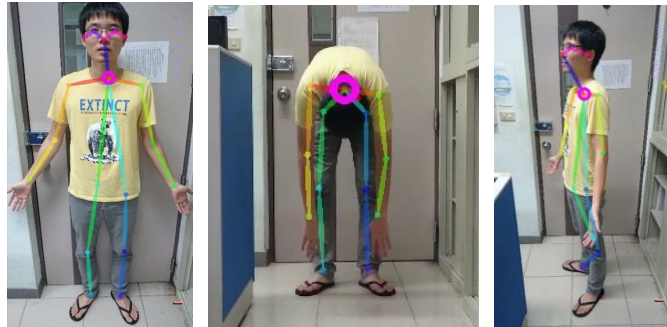
系統。希望此系統具備以下三個特點：(1) 易於架設的動作辨識系統、(2) 可利用彩色攝影機偵測動作及 (3) 提供客觀的日常生活動作評估。探討是採用人工選取的特徵，然後，搭配第一年所開發之 Deep JMLM 來加以辨識生活動作。此外，也會採用深度學習中的卷積類神經網路 (Convolutional Neuron Networks, CNN) 與之比較。

預計透過以下步驟來建構一個居家生活動作辨識系統：

步驟一：彩色影像輸入。

步驟二：擷取身體骨架資訊。

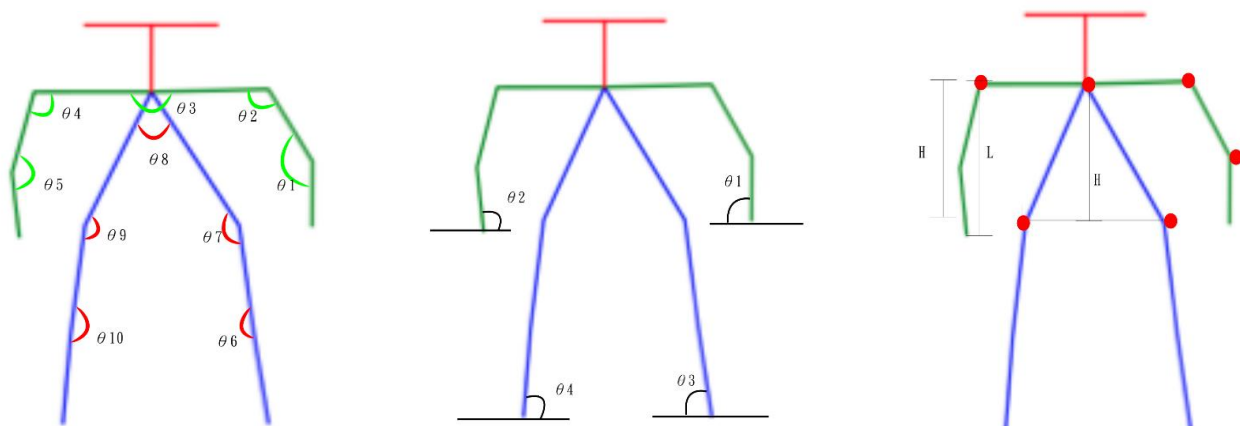
預計透過參考文獻的 [29] 中的骨架擷取方式來從彩色影像中擷取身體骨架資訊 (如圖五所示)。



圖五、彩色影像中擷取身體骨架資訊

步驟三：特徵擷取。

特徵的截取對於類神經網路而言是非常重要的，選取到不好的特徵會使神經網路的收斂較慢，甚至沒有辦法做有效的辨識；然而良好的特徵則相反，其能夠使類神經網路的收斂速度加速之外，辨識率也會有顯著的提升。我們認為身體的姿態 (如：站立、坐姿及彎腰等) 都會與關節角度及手腕相對應關節點的長度有關，所以，初步會計算以下三大類特徵 (如圖六所示)。(1) **四肢關節角度**：動作的組成主要是上下肢的組合來做出不同的動作，所以在關節的選取上只取上下肢之角度，在上肢分別為左手肘、左肩、脖子 (於左、右肩之角度)、右肩及右手肘共五個；在下肢五個分別為左膝、左髖關節、脖子 (於左、右髖關節之角度)、右髖關節及右膝等共五個，總共 10 個關節角度。(2) **四肢末端與水平的夾角**：人體骨架會和所在的空間有關，例如在蹲著和站著的時候，人體的小腿相對於地面的角度就是一個很好的辨識依據。(3) **手腕相對應關節點的長度變化**：手腕的相對位置是評估動作的另一項指標，尤其是在判斷舉手、坐姿和蹲姿等的幾種動作時手腕的相對高度可以提供明確的判斷依據。



圖六、姿態的三類特徵

步驟四：收集資料。

收集各種真實居家生活的情境，透過多人多次的方式收集足夠的訓練資料和測試資料。

步驟五：Deep JMLM 的建立。

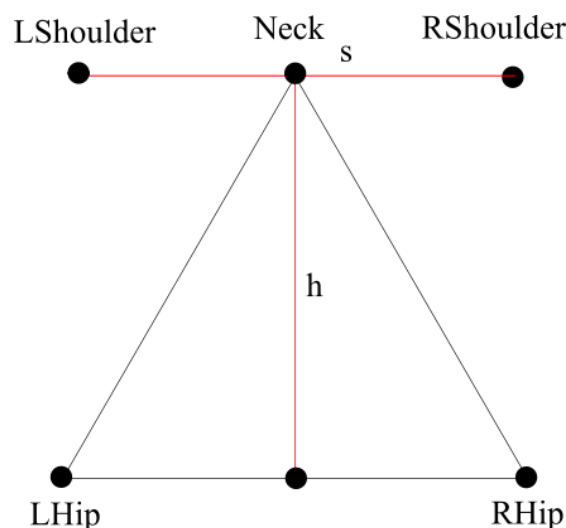
將上述的資料分為訓練和測試資料集，利用訓練資料集來建立 Deep JMLM，再用測試資料集來驗證辨識率是否有達到要求，否則要再重新訓練或尋找更有效之特徵。

步驟六：後處理。

本系統在辨識動作時所採取的動作是以分類每張影像來決定當下是做何類動作，但為了使系統更穩定避免掉雜訊的干擾，在後處理的時候會使用滑動窗口 (sliding window) 來慮除系統的雜訊，之後才能提供某段時間內的活動度量測報告。

1. **困難：**人所站立的位子並非是固定的，所以系統取得的骨架大小是很不一致的，而且每一個人的骨架資訊都會因為各人的高矮胖瘦而不相同，為了避免得到的骨架資訊差異太大，造成神經網路辨識上的困難，必須要將骨架資訊做相對應的正規化，從而產生更好的特徵以供類神經網路辨識。因此，如何正規化？

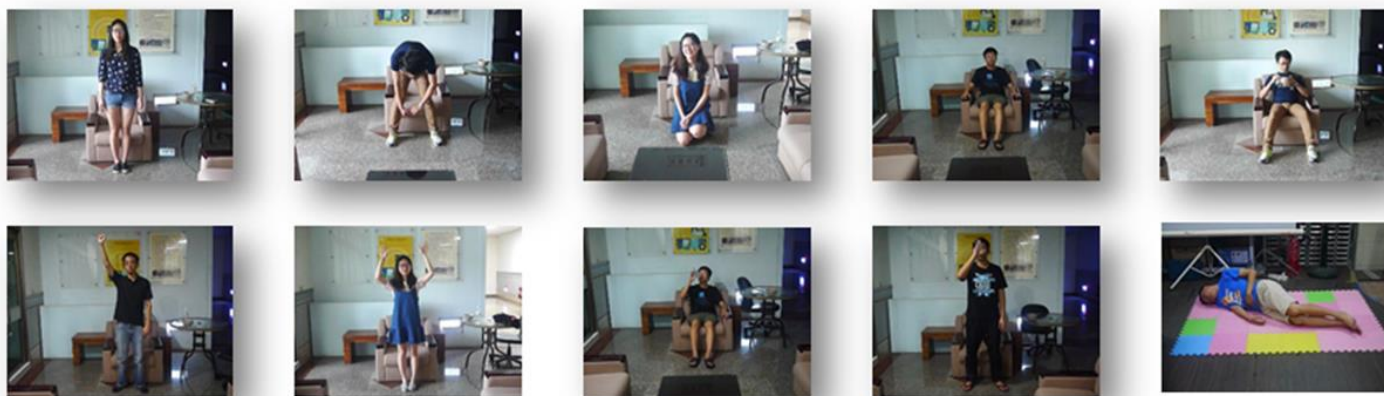
解決途徑：我們初步預計會採用兩步驟方式來正規化 (如圖七所示)。(1) 骨架平移：此部分是以脖子做為新的原點做平移，每一個關節點減去和脖子的坐標來做平移。(2) 骨架縮放，在二維的影像上抓到的骨架無法還原出原始的長度，只能點出骨架在二維影像上的坐標，無法得知實際肢體的長度，如果選取單一段肢體長度做為單位長度的話，則在攝影機與骨架成特定角度時該單位長度會失去其參考價值。所以，我們預計會以兩種肢體長度做為正規化時的依據。在做正規化時採取上述兩種不同單位長計算方式來做正規化，其中一種是以兩側髖關節中心到脖子的距離做為單位長度，另一種是以兩測肩關節的距離為單位長度，要依當時的骨架資訊來決定採用何種單位長度。



圖七、骨架參考單位長度示意圖

2. **困難：**如何選取該辨識多少種姿態來評估日常生活的活動度才具代表性？

解決途徑：經過初步的分析(如：表一)，我們預計會分類以下十種姿態 (如圖八所示)，因為各自有其代表意義。



圖八、十種姿態

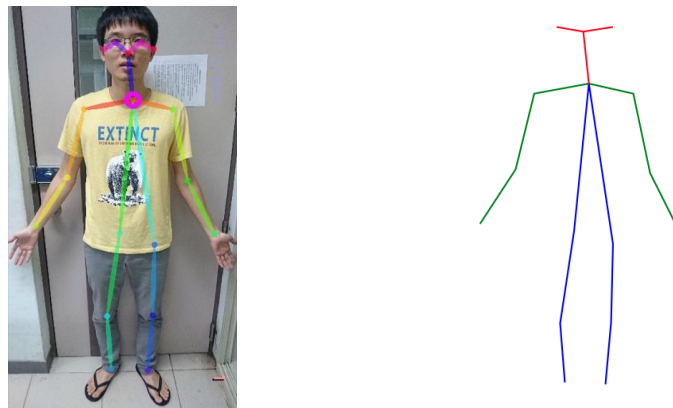
表一、十種姿態及其代表意義

代表姿態	代表意義
站立	站立在居家生活動作中是很普遍的一個動作，通常有幾個動作可能會有此骨架出現，如站立和走路等
彎腰	會有彎腰的骨架出現的時候，通常為撿拾地上物品時所產生此類骨架
蹲姿	蹲姿在居家生活中通常也為撿拾物品，只是以蹲的方式來撿拾
坐姿	在居家生活動作中會大有很多種坐姿出現，此為單純的坐著，且手部並沒有做其他的事情，通常是受試者於看電視或是休息時所呈現的動作
吃東西	吃東西的這個骨架動作為坐著且雙手於胸前的骨架動作，此類動作有吃東西、玩手機和看書等
舉單手	舉單手在通常在居家生活中拿高處的東西的會有此骨架動作出現
舉雙手	舉雙手通常為伸暖腰或者要拿用雙手拿東西時會有此動作出現

喝水坐	此動作為坐著時單手接近耳朵，會有此骨架的動作如坐著喝水和坐著接電話
喝水站	此動作為站著時單手接近耳朵，會有此骨架的動作如站著喝水和站著接電話
跌倒	在居家生活的情境中，本系統是以客廳為主，而在客廳中會有倒下的骨架發生時，通常為使用者跌倒時，做為簡單的跌倒偵測機制

3. 困難：如何與深度學習中的卷積類神經網路（Convolutional Neuron Networks, CNN）比較？

解決途徑：因為卷積類神經網路近幾年來被廣泛地應用於影像識別中，所以，我們也想與之作比較。那關鍵的問題是如何將何類影像輸入卷積類神經網路？我們預計會先將圖片的骨架資料拿到之後，將正規化過的骨架繪出作為卷積類神經網路的輸入資料，用以增加卷積類神經網路的辨識能力。在繪製骨架資料方面，先將正規化後的關節點分成三個部分，第一部分為頭部，分別為左眼、右眼及鼻子；第二部分為上肢，分別為左手腕、左手肘、左肩、脖子、右肩、右手肘及右手腕；第三部分為下肢，分別為左腳踝、左膝、左髖關節、右髖關節、右膝及右腳踝，總共 16 個關節點，其中各個肢體以不同的顏色作為表示，頭部為紅色；上肢為綠色；下肢為藍色。圖九顯示一個範例，其左側為原始圖片，右側為按照上述方式所做之輸出圖片。



圖九、卷積類神經網路輸入骨架影像

第三年：孩童的學習姿態辨識系統

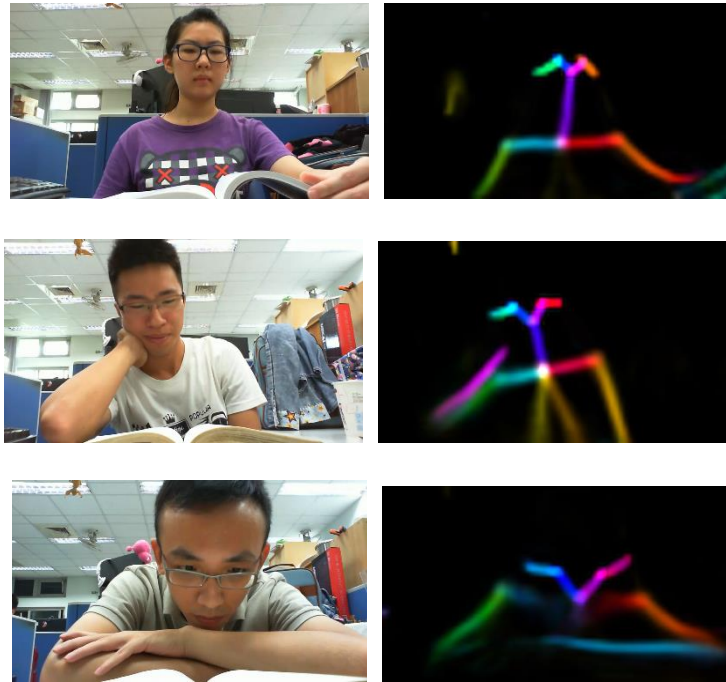
當父母親不能陪伴在小孩身邊時，父母親都關心小孩在自己的房間念書時，是否真的有專心在念書？還是東摸西摸地打發時間而已？抑或是趴在書桌上睡著了？若是小孩的確有認真念書，雖然會感到很欣慰，但也會擔心小孩是否有保持適當的閱讀距離？許多的父母親真的關心小孩的**專注力和視力問題**，所以，我們提出**孩童的學習姿態辨識系統**來陪伴孩童讀書，其目標是要減輕父母親的負擔和掌握孩童學習的狀態。

預計透過以下步驟來建構一個孩童的學習姿態辨識系統：

步驟一：彩色影像輸入。

步驟二：擷取身體骨架資訊。

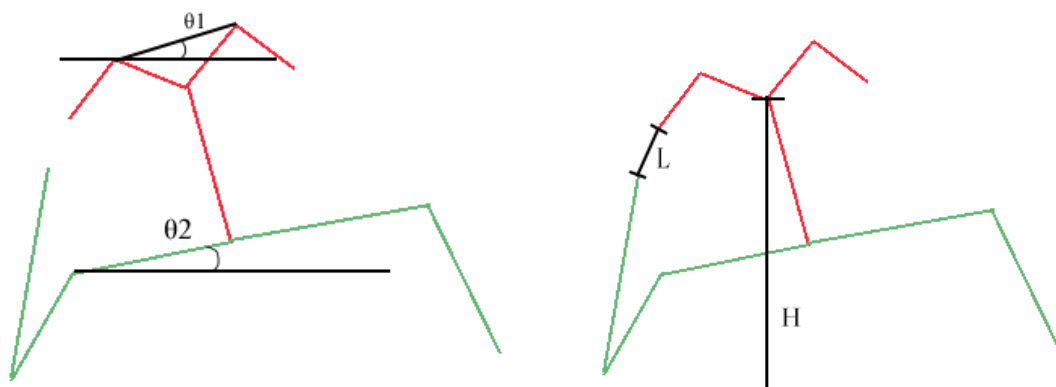
預計透過參考文獻的 [29] 中的骨架擷取方式來從彩色影像中擷取身體骨架資訊 (如圖十所示)。



圖十、閱讀姿態之身體骨架資訊

步驟三：特徵擷取。

特徵的截取對於類神經網路而言是非常重要的，選取到不好的特徵會使神經網路的收斂較慢，甚至沒有辦法做有效的辨識；然而良好的特徵則相反，其能夠使類神經網路的收斂速度加速之外，辨識率也會有顯著的提升。我們認為閱讀的姿態 (如：正常姿態、單手撐頭、歪頭和趴著等) 都會與關節角度、手腕與其他關節點的距離及關節點於影像中的位置有關。所以，初步會計算以下四種特徵 (如圖十一所示)。(1) **雙眼向量與水平的夾角**：臉部資訊是評估姿態的一項指標，主要可以辨識撐頭和歪頭的姿態。(2) **肩膀向量與水平的夾角**：在辨識左右稱頭時提供一個判斷的基準。(3) **手腕與對應耳朵的距離**：手腕的位置是評估姿態的另一項指標，在判斷左右稱頭的姿態時手腕與對應耳朵的距離可以提供明確的判斷依據。(4) **鼻子於影像中的位置**：在趴下時臉部資訊於影像中的位置會有明顯的改變。



圖十一、閱讀姿態的特徵

步驟四：收集資料。

收集各種真實孩童閱讀的情境，透過多人多次的方式收集足夠的訓練資料和測試資料。

步驟五：Deep JMLM 的建立。

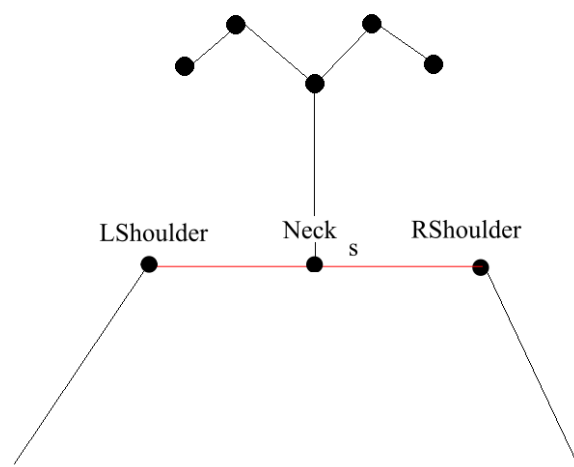
將上述的資料分為訓練和測試資料集，利用訓練資料集來建立 Deep JMLM，再用測試資料集來驗證辨識率是否有達到要求，否則要再重新訓練或尋找更有效之特徵。

步驟六：後處理。

本系統在辨識動作時所採取的動作是以分類每張影像來決定當下當做何類動作，但為了使系統更穩定避免掉雜訊的干擾，在後處理的時候會使用滑動窗口（sliding window）來濾除系統的雜訊，之後才能提供某段時間內的活動度報告。

1. **困難：**人閱讀的姿態並非是固定的，所以在系統取得的骨架大小是很不一致的，而且每一個人的骨架資訊都會因為各人的高矮胖瘦而不相同，相機擺放位置的遠近也會影響到骨架資訊，為了避免得到的骨架資訊差異太大，造成神經網路辨識上的困難，必須要將骨架資訊做相對應的正規化，從而產生更好的特徵以供類神經網路辨識。因此，如何正規化？

解決途徑：我們初步預計會採用兩步驟方式來正規化(如圖十二所示)。(1) 骨架平移：此部分是以脖子做為新的原點做平移 to 影像的中心點。(2) 骨架縮放，在二維的影像上抓到的骨架無法還原出原始的長度，只能點出骨架在二維影像上的坐標，無法得知實際肢體的長度，假設我們在時間 t 時以兩側肩關節的距離做為單位長度。



圖十二、骨架參考單位長度示意圖

2. **困難：**如何選取該辨識多少種姿態來評估孩童閱讀學習的專心程度？

解決途徑：經過初步的分析，我們預計會分類以下七種姿態（如圖十三所示）：正常、右歪頭、左歪頭、右撐頭、左撐頭、雙手撐頭和趴在桌子。主要為正常的閱讀姿態和六種常見的不良閱讀姿態。





圖十三、七種姿態

(三) 預期完成之工作項目及成果。請分年列述：1.預期完成之工作項目。2.對於參與之工作人員，預期可獲之訓練。3.預期完成之研究成果（如實務應用績效、期刊論文、研討會論文、專書、技術報告、專利或技術移轉等質與量之預期成果）。4.學術研究、國家發展及其他應用方面預期之貢獻。

第一年：

預期完成之工作項目：

- (1) 完成 Deep JMLM 的學習演算法。
- (2) 將Deep JMLM 應用於公用資料集的辨識，完成 Deep JMLM 與現有其他類神經網路的效能比較。
- (3) 論文撰寫。

參與之工作人員，預期可獲之訓練：

- (1) 學會類神經網路及深度學習的理論與技術。
- (2) 學會 Deep JMLM的設計。
- (3) 可提高計劃參加者之程式語言撰寫能力。。
- (4) 懂得跨領域之整合及溝通協調能力。

第二年：

預期完成之工作項目：

- (1) 完成生活姿態辨識的特徵擷取。
- (2) 完成居家生活動作辨識系統。
- (3) 論文撰寫。

參與之工作人員，預期可獲之訓練：

- (1) 可提高計劃參加者之程式語言撰寫能力。
- (2) 學會動作辨識技術。
- (4) 學會類神經網路及深度學習技術。
- (4) 學會跨領域之整合。
- (5) 可申請專利。
- (6) 論文撰寫能力。

第三年：

預期完成之工作項目：

- (1) 完成閱讀姿態的特徵擷取。
- (2) 完成孩童的學習姿態辨識系統。
- (3) 論文撰寫。

參與之工作人員，預期可獲之訓練：

- (1) 可提高計劃參加者之程式語言撰寫能力。
- (2) 學會動作辨識技術。
- (4) 學會類神經網路及深度學習技術。
- (4) 學會跨領域之整合。
- (5) 可申請專利。
- (6) 論文撰寫能力。

(四) 整合型研究計畫說明。如為整合型研究計畫請就以上各點分別說明與其他子計畫之相關性。