



基於資料擴增之行為辨識

陳文輝¹ 竇志浩²

1. 國立台北科技大學自動化研究所 教授
2. 國立台北科技大學自動化科技研究所 研究生

摘要

由於科技的進步，現在許多硬體設備有著非常好的運算能力，而深度學習也因此有著突破性的發展，成為目前熱門的研究與發展方向，因為使用深度學習架構，只需要足夠的資料量，和良好的網路架構，與傳統的機器學習方法相比，往往都有較好的結果。若深度學習模型缺乏資料量，則其效能就不會那麼理想。本文以行為辨識為基礎，結合感測器資料與深度學習方法，利用感測器資料去進行資料擴增改善辨識結果。由於一般的影像資料擴增方法，並不適用於感測器資料，所以我們親自蒐集部份資料進行實驗，並且提出了五種方法做資料擴增：混入雜訊、時間序列置換、感測器的資料旋轉以及兩種深度學習生成模型，使用少量的訓練資料，建立更高效能的模型。本研究所使用的方法針對 UCI 數據集其準確率可以達到 98.32%，與過去未使用資料擴增方式的行為辨識相同分類器比起來，最高可以提升 1.5%。

關鍵詞：資料擴增、生成對抗網路、變分自動編碼器、深度學習

*Corresponding author Email : t105618024@ntut.edu.tw



Data Augmentation for Human Activity Recognition

Wen-Hui Chen¹, Chih-Hao Tou²

1. Professor, National Taipei University of Technology, Graduate Institute of
Automation Technology

2. Graduate Student, National Taipei University of Technology, Graduate Institute of
Automation Technology

ABSTRACT

Due to the advances in technology and hardware development, deep learning has become a popular research topic in recent years. In comparison with traditional machine learning algorithms, deep learning based approaches have better performance in many tasks. However, the application of deep learning requires ample data and a well-designed neural network architecture. When the training data is not enough or limited, we do not normally obtain good results using deep learning approaches. This thesis proposed an approach based on the fusion of deep learning and data augmentation to improve the activity recognition accuracy. We proposed five methods for data augmentation, including noise injection, permutation, rotation, and two different generative models to reach a high-performance model with limited training data. The experimental results showed that the proposed data augmentation approach could reach an accuracy of 98.32% on a UCI dataset, and boost the accuracy up to 1.5% when compared with data without augmentation.

Keywords : Data augmentation, GAN, VAE, Deep learning, Activity Recognition

*Corresponding author Email : t105618024@ntut.edu.tw

一、前言

科技進步造成手持式裝置與穿戴式裝置的發展迅速且多樣化，所囊括的功能也是日益漸增，現代人日常生活中已經無法與這些行動裝置分離，然而在行動裝置內的感測器功能也是越來越複雜，像是加速度計、陀螺儀、方向感測器...等。在過去進行行為辨識時，身上許多部位都必須穿戴許多的感測器，而感測器又會因為精度的不同，價格也跟著不同，但是現在只需要一部行動裝置，透過簡單的幾個指令，就可以完成過去需要穿戴許多感測器的工作，在資料整理與傳輸上還更加方便，價格也相對低廉。雖然行動裝置帶給了蒐集資料有許多的便利性，但是在資料蒐集還是潛藏著一個嚴重的問題即資料量不足。資料蒐集除了單純的收集感測器資料之外，這些資料也需要有大量的人力以及時間去分類標籤，在這樣的狀況下使得可以使用的資料取得更加困難。

現在有許多深度生成模型，像是生成對抗網路(generative adversarial network, GAN)與變分自動編碼器模型(variational autoencoder, VAE)，這些模型在許多不同的領域，都取得了非常良好的結果，比如：逼真的高分辨率影像(Martin Arjovsky, 2017)、來自文本描述的真实影像(Emily L Denton, 2015)，甚至包括音樂作品(Li-Chia Yang, 2017)和語音(Heiga Zen, 2017)，以高分辨率影像來說，可以產生出與訓練資料相似，但是又不完全一樣的影像，這樣可以大大的解決蒐集資料的問題。但是在感測器的應用上，卻沒有那麼好的結果，在神經網路中，資料量的多寡往往可以決定模型好的壞，所以在資料量缺乏的狀況下神經網路模型的結果並不會那麼好，本研究中將利用傳統方法與深度學習(deep learning)的資料擴增(data augmentation)方法，解決模型在資料量缺乏的狀況下效果不是非常理想的問題。

本研究使針對 UCI 公開資料庫的資料集(Jorge-L, 2015)，由加速度計與陀螺儀共六軸資料且含有 6 種不同的行為，利用資料擴增的方法改善模型的準確度。本研究目的為：(1) 有效增加資料量，改善模型效能 (2) 比較傳統機器學習與深度學習方法的優劣 (3) 實驗基於整體學習的深度學習分類器並且應用於行為辨識中。

二、資料擴增

2-1 傳統資料擴增

傳統的資料擴增，大部分都是應用在影像中，並且基於幾何原理來做變換，有放大、縮小、旋轉、反向等等許多不同的方法，但不是每種方法都適用在感測器資料，所以我們在提出以下這幾種擴增方法用於感測器資料，混入雜訊(noise injection)、時間序列置換(time series permutation)和旋轉(rotation)，這些改變都是屬於細微的變化，並不會改變數據的標籤，改變出來的數據都是非常有可能發生在生活中的。感測器的數據常常會因為感測器所擺放的位置而有數據的變化，就算是同一個標籤中，同樣的位置，只要螢幕面向的方向或是裝置放置的方式不同，都會造成最後所取得的數據有所差別，為了解決螢幕面向的方向與裝置放置方式的問題，我們可以使用旋轉擴增資料的方式，或是只使用符號的轉換方式解決面朝的問題。每個人的行為模式不一樣，以走路為例，不同的人走路速度不一樣，對於走快與走慢所蒐集到的數據也會因此不同，只是有些許的偏差，所以可以加入隨機的高斯雜訊，擴增出不同走路速度的資料，也可以透過時間序列置換的方式，將一個窗口中的資料分成 N 等分的資料在進行隨機互換位置，去增加不同走路行為的資料。

2-1-1 混入雜訊

混入雜訊是一種可以增加資料又不會影響到標籤的辦法，以感測器數據來說，不同的人在做相同的動作時，重心與力道都不會一樣，造成數據上會有些許的不同，所以我們利用這樣子的關係，給數據集中的數據添加一些隨機雜訊，讓原本的資料增加，而且不會影響到數據的標籤。高斯雜訊是一種在影像中常使用混入雜訊的方式，高斯雜訊是以符合高斯分布(常態分佈，如圖 1)的機率密度函數所生成的如圖 2 所示，高斯分布如式(1)：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (1)$$

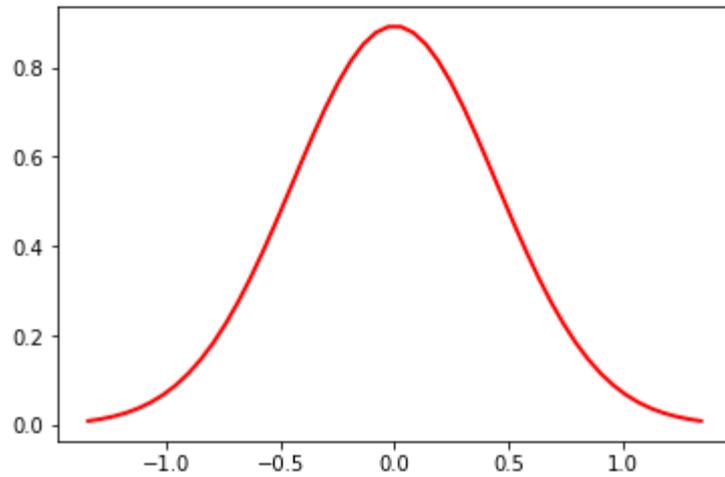


圖 1 高斯分布圖

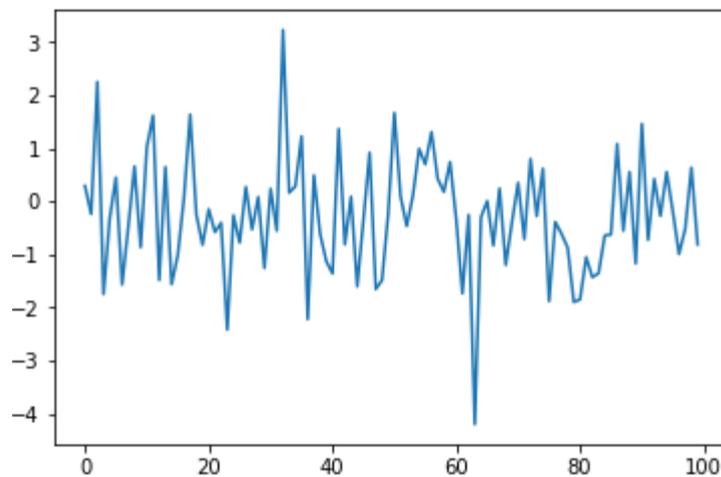


圖 2 高斯雜訊示意圖

2-1-2 時間序列置換

時間序列置換是指將在某段時間內的資料互相交換位置，在序列資料中，每一筆資料都是包含一定的時間，所以我們可以利用一段時間的動作，將其拆分為同等份許多段的資料，再將其重新排列組合，由於是經由更改時間順序的排列組合，並不會影響到標籤，若以原始資料的角度來看，時間序列置換是沒有意義的，因為只是交換資料的排序，但是對於含有時間步階的資料來說時間序列置換的影響是非常大的。

2-1-3 旋轉

在蒐集資料的過程中，感測器往往都必須以固定的位置擺放，避免不同的位置會造

成感測器數據的誤差，然而在此我們利用三維旋轉的原理，將三軸的資料以不同的角度去做旋轉，這樣子所產生出來的資料，可以更加貼近生活習慣，以手持行動裝置為例，有些人習慣橫著用，有些人習慣豎著用，這些不同的習慣，往往會造成感測器的資料數值不同，但是卻是在做同一件事情，所以我們利用旋轉，產生出不同角度的資料，來幫助模型學習到更多資訊。

由於加速度計與陀螺儀所取得的資料為 x, y, z 三軸的資料，所以在此我們將加速計與陀螺儀分開處理，皆使用三度空間的旋轉方式進行資料的旋轉。手機的感測器在平放時，螢幕的方向為 z 軸，裝置的左右邊為 x 軸，裝置的上下為 y 軸，如圖 3 所示。我們可以將手機的平放、側躺、斜靠等，不同的手持方式放置方式，想像為在三維空間中的旋轉，所以利用這個原理，我們將平放的手機感測器資料，去做三維空間的旋轉，給定旋轉的角度與基於哪軸旋轉，求得不同放置方式，所取得出的數值。

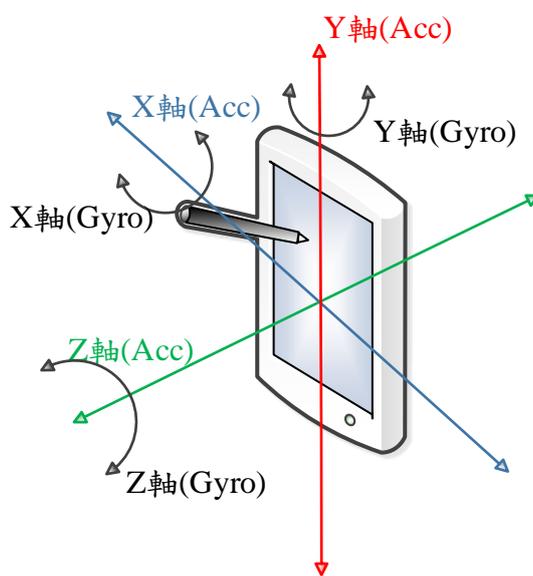


圖 3 手機加速度計與陀螺儀三軸示意圖

依照尤拉旋轉定理(Euler's rotation theorem)，在三維空間中，假設一個物體在位移的時候，物體內部至少會有一個固定點，此位移等於一個繞著包含那固定點的固定軸的旋轉。在三維空間的旋轉中，任何旋轉都可以用三維空間中的 x 軸、 y 軸與 z 軸的旋轉角

度來表示，而這樣的表示又可以分為兩個方法，一種是三軸是固定的並不會跟著旋轉而旋轉，稱為外旋轉(extrinsic rotation)；另一種是三軸是非固定的並會跟著旋轉而旋轉，稱為內旋轉(intrinsic rotation)，在此我們所使用的是外旋轉。假設 x 軸、 y 軸與 z 軸的旋轉角度分別為 $\theta_x, \theta_y, \theta_z$ ，則旋轉矩陣可以表達如下：

$$R = R_y(\theta_y)R_x(\theta_x)R_z(\theta_z) \quad (2)$$

$$R_x(\theta_x) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta_x & -\sin \theta_x \\ 0 & \sin \theta_x & \cos \theta_x \end{bmatrix} \quad (3)$$

$$R_y(\theta_y) = \begin{bmatrix} \cos \theta_y & 0 & \sin \theta_y \\ 0 & 1 & 0 \\ -\sin \theta_y & 0 & \cos \theta_y \end{bmatrix} \quad (4)$$

$$R_z(\theta_z) = \begin{bmatrix} \cos \theta_z & -\sin \theta_z & 0 \\ \sin \theta_z & \cos \theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (5)$$

2-2 深度學習資料擴增

機器學習領域中，可以將所有的方法歸類為兩大類，分別為生成模型(generative model)與判別模型(discriminative model)，生成模型是利用機率統計的方法去計算觀測數據的資料分布，以一個隱函數去擬合觀測數據的資料分布，而後生成假的觀測數據。而判別模型則是，學習資料的特徵，並將其作分類或是預測。以下將介紹生成對抗網路與變分自動編碼器模型這兩種不同的深度生成模型。

2-2-1 變分自動編碼器模型

自動編碼器(autoencoder, AE)是一種非監督式的學習方法，基本上由三層神經網路所組成，第一層是編碼層(encoder layer)，編碼層的神經元數量小於輸入資料的神經元數量，利用神經元數量較少與激勵函數(activation function)來達到資料壓縮的效果，第二層為隱藏層(latent layer)，隱藏層是經過編碼層後達到神經元數量最小的狀態，第三層為解碼層(decoder layer)，解碼層就是將編碼過後的隱藏層，重新解碼為與輸入資料神經元數量相同的資料，層與層之間都是採用全連接(fully connected)的概念進行操作。自動編碼器的作用是將輸入的資料透過全連接層，壓縮為較低為度的隱藏層，再利用解碼層解壓

縮回原本的資料，但是在壓縮的過程中，會不可避免的丟失一些信息，所以解碼後的資料，會產生一些雜訊，圖 4 為自動編碼器示意圖。

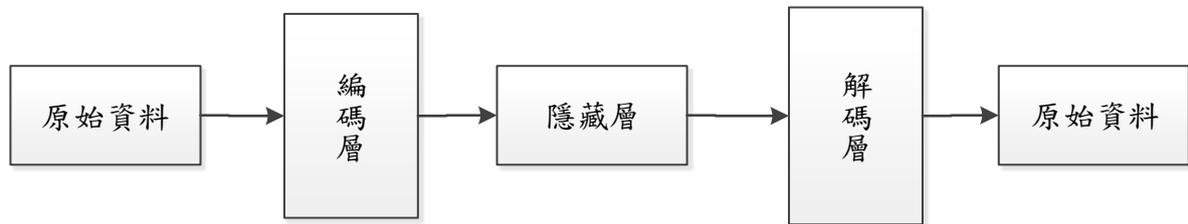


圖 4 自動編碼器架構圖

變分自動編碼器(Kingma, 2013)與自動編碼器不同在於，他將機率的模型加入了隱藏層中，在變分自動編碼器的隱藏層中，他所產生的兩個向量，一個是均值向量，一個是標準差向量，利用這兩個向量，與高斯常態分布結合，產生出新的資料隱藏分布，再將新的資料隱藏分布，經過解碼層，解碼層新的資料。在變分自動編碼器的損失函數(loss function)的部分，一方面我們與自動編碼器相同，使用平均平方誤差來度量原始資料與新資料的差異，另一方面我們也使用了相對熵(Kullback Leibler divergence, KLD)作為隱藏分布與高斯分布的度量，KLD 是用來計算兩不同機率分布差距的算法，在 VAE 中由於我們需要將新生成的樣本資料分布與原始樣本資料分布做比較，所以在此我們選用 KLD 算法來當作度量。圖 5 為變分自動編碼器示意圖。

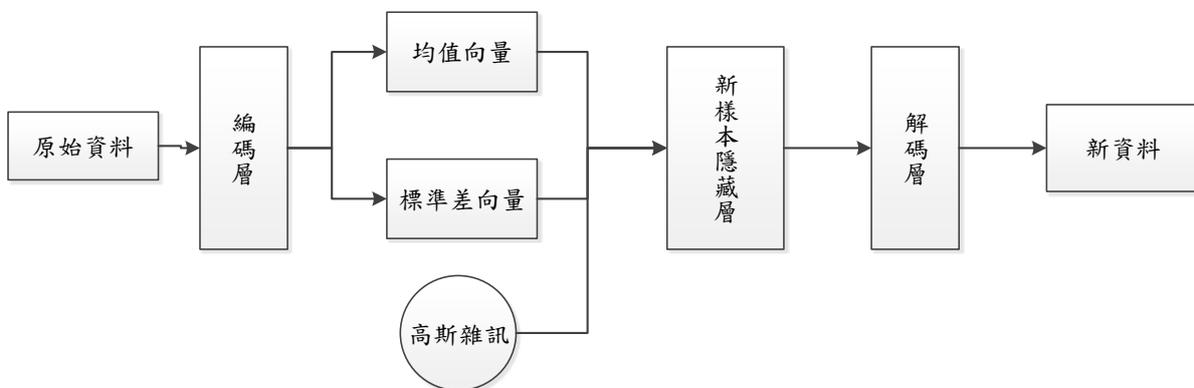


圖 5 變分自動編碼器架構圖

2-2-2 生成對抗網路

生成對抗網路的基本想法是來自於賽局理論(game theory)，由兩個人所組成的遊戲，這兩個人可以稱為生成網路(generator network)與判別網路(discriminator network)。生成網路的作用就是生成我們所需要的東西，判別網路的作用就是不斷的去判別生成網路的作品，直到生成網路能夠做出我們所要的東西，而判別網路也分不出來這樣東西是真實的還是生成網路所製造出來的虛擬的東西，如圖 6。

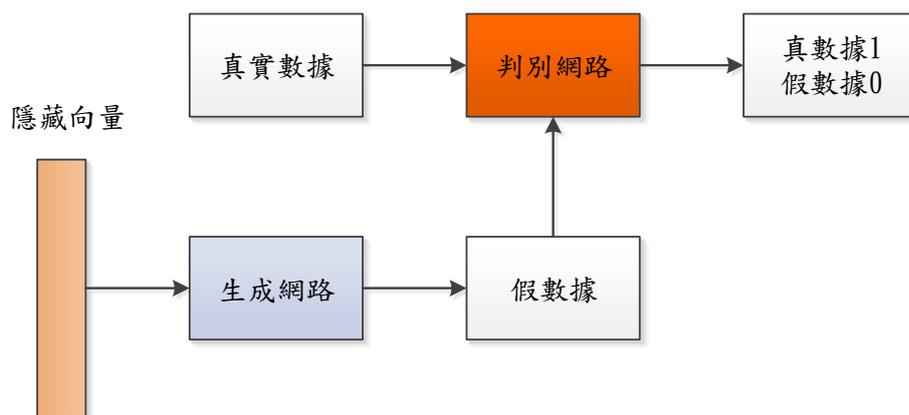


圖 6 生成對抗網路示意圖

在做生成對抗網路的訓練時，我們必須先對判別網路進行訓練，如圖 7。讓判別網路先學習到何謂真實數據，再將生成網路生成的數據輸入到判別網路中，讓判別網路學習何謂假冒數據，最後我們將判別網路的參數固定後，去更新生成網路的參數，讓經過生成網路後生成的資料能夠被判別網路判別為真實數據，這樣我們就完成了生成對抗網路的訓練。

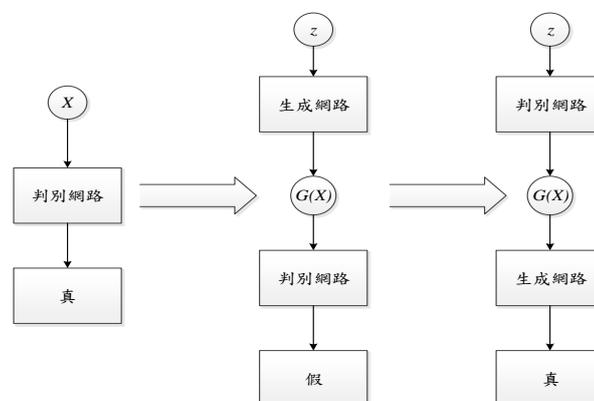


圖 7 生成對抗網路訓練流程圖

三、結果與討論

本章詳細解釋了本研究的實驗方法，論文中所使用的硬體規格與開發工具清單詳見表 1，實驗流程詳見圖 8。在這次的實驗中我們使用了 UCI 公開資料庫的資料進行實驗驗證與分析。本次實驗探討不同資料增加的處理下，對相同模型的影響，對原始資料模型做比對，接著再更改不同的分類模型並比較結果，最後將本文的方法與國內國外之相關行為辨識的文獻做比較。

表 1 硬體規格與開發工具清單

硬體規格與開發工具	
作業系統	Microsoft Windows 10 專業版 64 位元
處理器(CPU)	Intel®Core™i5-6500 3.20GHz
記憶體(RAM)	2133 DDR4 8G
顯示卡(GPU)	ASUS GeForce GTX 1060 O6G
程式開發工具	Python3.6 Keras2.1.0 Tensorflow 1.7

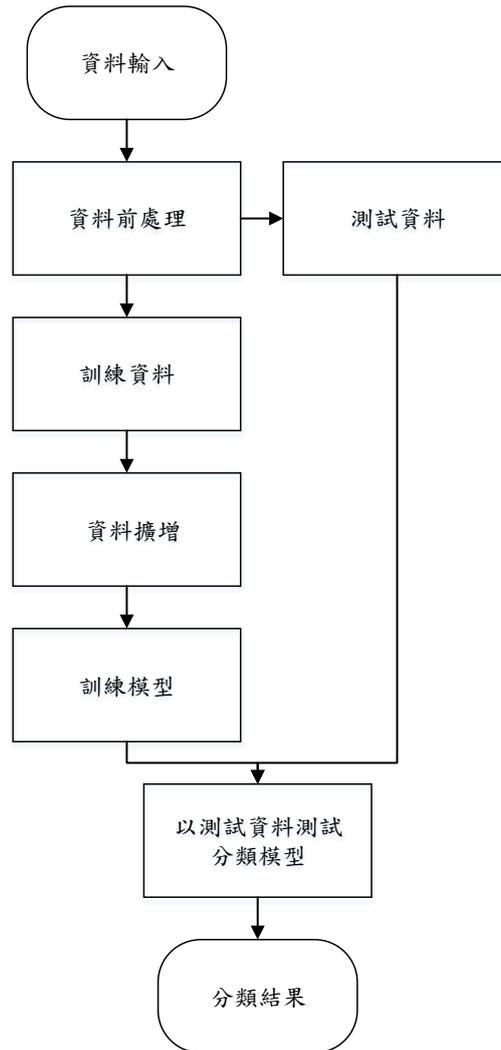


圖 8 實驗流程圖

本次實驗我們採用了 UCI 所提供的公開資料庫，其資料為 30 名自願者，年齡區間為 19-48 歲，實驗過程中，每個人在腰部配戴智慧型手機(Samsung Galaxy S2)進行三項靜態姿勢(坐、站立、躺下)和三項動態姿勢(步行、上樓梯、下樓梯)。實驗數據包括姿態的轉換。該資料庫使用智慧型手機內建的加速度計與陀螺儀，以 50Hz 的取樣頻率，去獲取 3 軸直線加速度與 3 軸角速度的數據，實驗過程中有拍攝視頻佐證，並且用手動標記數據。由於取樣頻率為 50Hz，所以一秒有 50 筆資料，通常在判斷這個人是否在做這個行為時，通常要有 2-3 秒的時間，所以我們把時間步階設定為 100，每筆資料就包含著兩秒的資料量。覆蓋率(overlap)設定為 0，表示每筆資料互相獨立並沒有關係。

表 2 UCI 資料各個行為資料數量

資料型態 行為	原始	時間步階=100
步行	122091	1220
上樓梯	116707	1164
下樓梯	107961	1076
坐	126677	1264
站立	138105	1380
躺下	136865	1366
總共	74840	7470

3-1-1 資料增加結果

將各種不同的資料擴增方法應用在 UCI 資料集中，得到了經過各種不同方法擴增資料後的訓練資料，在此測試資料是一開始就先隔開的，所以並不會將測試資料進行擴增，擴增後的各個方法訓練與測試資料比例如下表 3。

表 3 經過傳統擴增方法後各訓練資料與測試資料的各個標籤數量與比例

標籤	訓練資料 (noise)	訓練資料(permutation)	訓練資料(rotation)	測試資料
步行	1952(16.31%)	1952(16.32%)	1952(16.31%)	244(16.37%)
上樓梯	1886(15.61%)	1864(15.59%)	1886(15.61%)	232(15.59%)
下樓梯	1726(14.42%)	1724(14.42%)	1726(14.42%)	214(14.36%)
坐著	2026(16.93%)	2024(16.92%)	2026(16.93%)	252(16.91%)
站著	2208(18.45%)	2208(18.46%)	2208(18.45%)	276(18.52%)
躺著	2188(18.28%)	2188(18.29%)	2188(18.28%)	272(18.25%)
總計	11966	11960	11966	1490

在變分自動編碼器模型中，我們將以六軸資料為輸入，經過兩層的編碼層後輸入到隱藏層中，隱藏層的神經元數量設定為 2 個神經元，在經過加入高斯雜訊後，變成新的樣本，最後再輸入到解碼層中，將資料還原。變分自動編碼器模型參數如下表 4，表 4 的參數皆為考量到電腦效能與進行多次實驗後得到較好結果的參數。加入生成資料後訓練資料與測試資料的各個標籤比例如下表 5。

表 4 變分自動編碼器模型參數表

參數	變分自動編碼器
批量大小(Batch Size)	3000
輸入維度(Input Dimension)	6
解碼層維度(Encode Dimension)	3
隱藏層維度(Latent Dimension)	2

表 5 加入變分自動編碼器後訓練資料與測試資料的各個標籤數量與比例

標籤	訓練資料(比例)	測試資料(比例)
步行	1952(16.35%)	244(16.37%)
上樓梯	1864(15.59%)	232(15.59%)
下樓梯	1724(14.43%)	214(14.36%)
坐著	2024(16.93%)	252(16.91%)
站著	2208(18.47%)	276(18.52%)
躺著	2178(18.23%)	272(18.25%)
總計	11950	1490

最後是經過生成對抗網路所生成的數據，在生成對抗網路中，我們將以六軸資料為目標，生成網路的目的就是要生成出辨識網路無法分辨出真假的六軸資料，而辨識網路就是將這六軸資料輸入去作辨識，生成對抗網路的參數如下表 6，表 6 的參數皆為考量到電腦效能與進行多次實驗後得到較好結果的參數。加入生成資料後訓練資料與測試資料的各個標籤比例如下表 7。

表 6 生成對抗網路參數表

參數	生成對抗網路
批量大小	3000
輸入維度	6
隱藏層維度	3
高斯雜訊維度(Noise Dimension)	100

表 7 加入生成對抗網路後訓練資料與測試資料的各個標籤數量與比例

標籤	訓練資料(比例)	測試資料(比例)
步行	3906(18.75%)	244(16.37%)
上樓梯	3734(17.92%)	232(15.59%)
下樓梯	3454(16.58%)	214(14.36%)
坐著	3140(15.07%)	252(16.91%)
站著	3314(15.90%)	276(18.52%)
躺著	3284(15.76%)	272(18.25%)
總計	20832	1490

最後將上述實驗中所使用的所有資料擴增方法結果，與原本的未經過任何處理的資料進行比較。比較時我們使用了四種不同的分類器，第一種為深度學習分類器，第二種為基於整體學習的深度學習分類器，第三種為決策樹，第四種為隨機森林，結果如圖 9。在此我們不只使用 UCI 的數據，也自行蒐集了六個不同動作的六軸資料進行比較，結果如圖 10。由圖中可以看出在資料擴增時，對於模型的辨識能力是有著良好的效益，而深度學習資料擴增的效果高於傳統資料擴增的方法，最高可以達到 98.32% 的辨識率。圖 9 與圖 10 中可以看到各個資料擴增方法對不同資料集的改善各不相同，可能是因為在資料擴增的方法上有所差異而造成的，而各個資料擴增方法對模型的改善各不相同，可能是因為使用資料擴增後的資料所抽取到的特徵不同而造成差異。

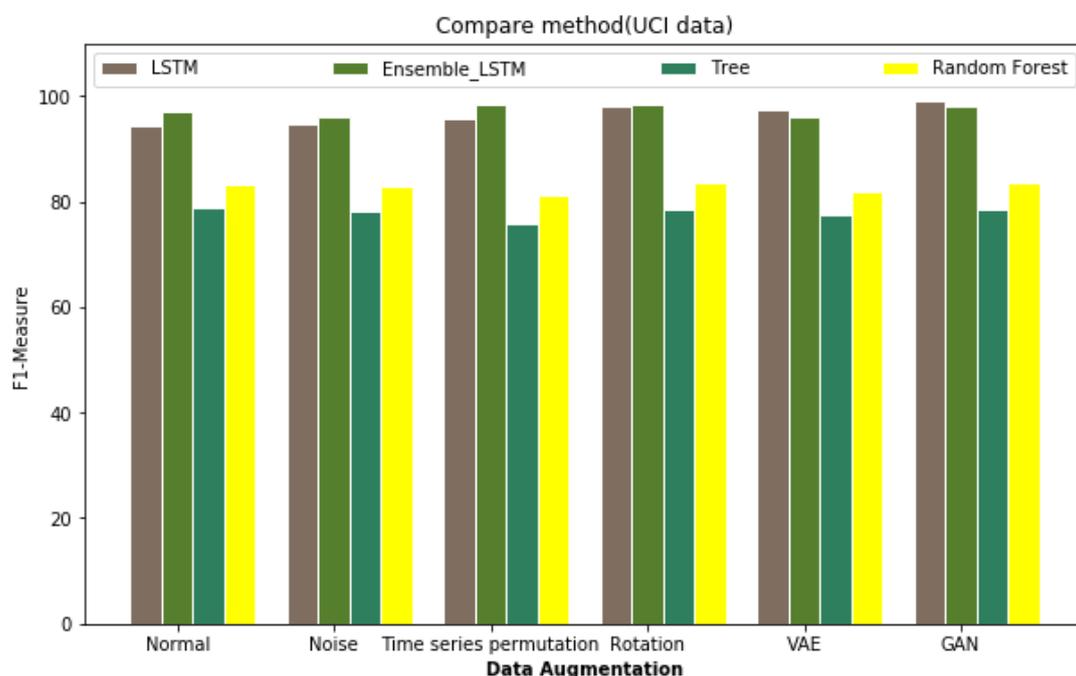


圖 9 使用 UCI 資料集各種擴增方法辨識率比較

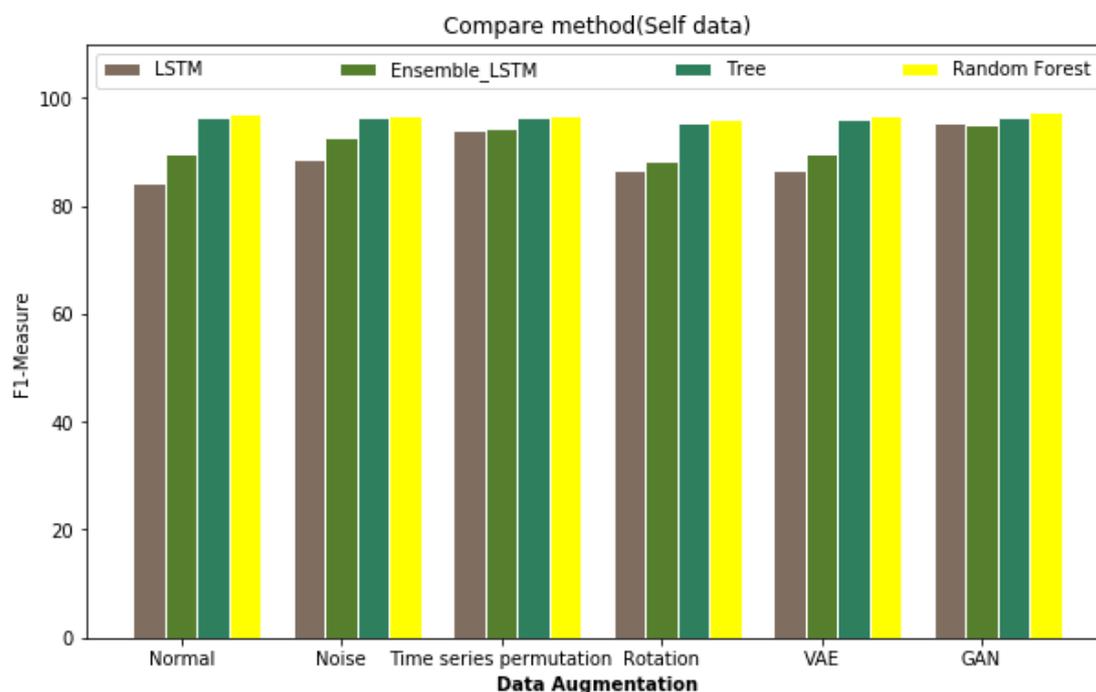


圖 10 使用自行蒐集資料的各種擴增方法辨識率比較

我們也與其他相關文獻進行比較，C.A.Ronao 等人(C.A.Ronao, 2015)利用深度卷積神經網路去處理一維信號的時間序列，而在深度卷積神經網路中，不僅利用了時間序列資料中固有的時間局部依賴性，以及動作平移不變性和分層特性，還使用深度卷積神經網路中自動提取特徵的方法，這樣可以免去繁雜的預處理，以減少人工消耗。在 C.A.Ronao 等人的研究中，使用多種不同深度的卷積神經網路，並且證明每增加一層都會產生更複雜的特徵，但隨著在卷積層中的信息傳遞，相鄰的層之間複雜度會減小，也證明大尺寸的濾波器對於本研究是非常有益的，因為讓相鄰的感測器讀數之間的時間局部依賴性具有更寬的持續時間，但是在池化層中則是越小越好，因為要維護數據的信息以免丟失重要資訊。Almaslukh 等人(Almaslukh, 2017)在研究中使用了堆疊自動編碼器 (stacked autoencoder, SAE)，在研究結果他們表明，他們使用堆疊自動編碼器的方法在每個樣本平均辨識的時間花費由 0.2724ms 減少到 0.0375ms，並與 UCI 數據集做比較，發現在模型效能上也是高於其他研究方法。本研究中我們不僅結合了變分自動編碼器模型與長短期記憶網路，而且在相同的數據集中的效能也是高於這些研究。詳見下表 8。

表 8 相關文獻之比較

研究	方法	資料集	模型效能
C.A.Ronao et al.,2015	Deep Convolutional Neural Network	UCI	95.75%
Almaslukh et al.,2017	Stacked Autoencoder	UCI	97.50%
本研究	GAN+LSTM	UCI	98.32%

四、結論

行為辨識在未來的智慧生活中佔據著一部分，在智慧生活中的居家系統能夠應用行為辨識系統，來判斷事件的輕重緩急，而設定優先處理順序。本研究將影像領域中的資料擴增方法應用在行為辨識中，可以有效地增加有限的資料。而本文也使用了幾種不同的分類演算法。首先，先讓資料經過資料擴增處理，將新增加的資料與原本的資料做結合，接者再透過傳統的機器學習演算法與深度學習對樣本建模，並且比較四種不同演算法的差異，本研究顯示經過資料擴增處理後的模型效能遠超過未經資料處理的模型效能，原始的模型辨識率為 95.5%，而經過資料擴增後的模型可達到 98.32%。由上述實驗結果可證實，經過資料擴增後的資料，不僅能解決資料缺乏的情況，還能夠建構出較完美的模型。本研究期望未來能夠減輕蒐集醫療資料的困難，能夠使擴增後的資料應用年齡層更加廣泛，並且在年長者的居家照護中提供正確的資料，幫助醫療人員進行資訊判斷，對使用者進行即時的治療與防護。

參考文獻

- Martin Arjovsky, Soumith Chintala, and Léon Bottou, 2017, “Wasserstein GAN” arXiv preprint arXiv:1701.07875.
- Emily L Denton, Soumith Chintala, Rob Fergus, et al, 2015 ,Deep generative image models using a Laplacian pyramid of adversarial networks. In Advances in Neural Information Processing Systems (NIPS), pp. 1486–1494.
- Li-Chia Yang, Szu-Yu Chou, and Yi-Hsuan Yang, 2017, “MidiNet: A convolutional generative adversarial network for symbolic-domain music generation using 1d and 2d conditions,” arXiv preprint arXiv: 1703.10847.
- Heiga Zen, 2017, “Generative model-based text-to-speech synthesis,” Invited talk given at CBMM workshop on speech representation, perception and recognition.
- Jorge-L. Reyes-Ortiz, Luca Oneto, Albert Samà , Xavier Parra, Davide Anguita, 2015 , “Transition-Aware Human Activity Recognition Using Smartphones.” Neurocomputing. Springer.
- Kingma, Diederik P and Welling, Max. “Auto-Encoding Variational Bayes,” In the 2nd International Conference on Learning Representations (ICLR), 2013.
- C. A. Ronao, and S. Cho, 2015 “Deep convolutional neural networks for human activity recognition with smartphone sensors,” in International Conference on Neural Information Processing, pp.46-53, Springer International Publishing.
- Almaslakh, B., AlMuhtadi, J., Artoli, A., 2017. An effective deep autoencoder approach for online smartphone-based human activity recognition. International Journal of Computer Science and Network Security 17, 160.