

智慧型汽車導論 期末報告-Survey of Driver Monitor System

成員:R11944037 張宏洋 R11922131 周柏諺 R11944050 洪郡辰
R11944058 何樂為

1. Abstract

本報告著重在多篇論文的survey, 對於各式駕駛監測系統, 從廣度層面的探討, 透過比較不同的角度來切入對於駕駛的偵測, 我們詳細的survey了6 篇論文, 將它們歸入 5 個類別, 並在本報告中進行介紹。

本報告在本學期之前沒有當作雙重作業來繳交且不為任何已完成的作業。

2. Introduction

不良的駕駛行為占事故原因的一大部分, 其中疲勞駕駛更是比正常駕駛高出四到六倍的事務風險, 因此, 駕駛監控系統的重要性不言而喻, 有一個安全、全面且舒適的駕駛監控系統是不可或缺的, 其中對於疲勞駕駛的偵測更是重要, 我們需要不管任何環境都能偵測出駕駛進入昏睡的狀態, 從各方面無死角的監測出駕駛的昏睡行為是我們這次survey的重點, 畢竟與車禍風險有關, 寧可多測也不能測不到。

往往系統對警報如何去通知駕駛通常是一筆帶過, 但我們對於如何根據駕駛的情況發出警報, 才能有效地把訊息傳給駕駛, 也有去關心如何解決。

3. Previous work or related work

傳統上, 駕駛監測主要使用相機, 近年來透過深度學習對駕駛行為進行分析取得了一定的成果, 因此我們選擇對這個方向進行研究。在閱讀了相關的論文後, 認為在real-world實現需要滿足以下三個條件: 1. 模型具備足夠的robustness。2. 模型能夠滿足汽車嚴格的real-time條件。3. 模型具備足夠的辨識能力。在滿足real-time及足夠的辨識能力是駕駛監測的一大挑戰, 必須在這之間做trade off。

在大多數透過深度學習對駕駛行為分析的論文其影像來源都源自於單一相機, 相機可能架設於側面或是正面進行拍攝。單一影像資料來源代表模型的robustness較低, 十分依賴於相機的狀況。如果相機鏡頭模糊影像傳輸延遲、相機故障等情況發生時

，駕駛檢測系統會馬上失效。因此多個影像資料來源是有必要的，多個影像資料來源不僅能夠提升模型的robustness也能提升模型對行為分析的準確率，因為有了來自不同角度的影像也代表著有了更多的特徵，這使得模型能夠有更多資訊進行行為判斷。

模型將駕駛行為分析視為一個多類別分類問題，將行為分為九類，分別為Normal Driving、Texting、Eating、Talking、Searching、Drinking、Watching video、Gaming、Preparing。在第二次實驗時將相似的行為結合，行為分為五類，分別為Normal Driving、Using Phone、Eat and Drink、Talking、Preparing。

在資料集方面，將相機架設於正面以側面拍攝，使用兩台相機的影像以及兩組影像的Optical Flow作為模型的輸入，以達到多影像來源的效果。

因為需要在任何情況都要能偵測出疲勞駕駛，我們也survey了幾種若相機不好偵測的情況下，基於駕駛生理現象依然能準確測出疲勞駕駛的方法，像是透過眼電圖及腦電圖的方法，眼電圖主要是因為角膜與視網膜之間存在電位差，故透過測量眼睛周圍的電位變化，來擷取眼睛移動的軌跡及眼部的動作，不過因為擷取到的訊號很小，所以擷取到的水平訊號差值及垂直訊號差值會先經由放大器放大，並且由於EOG是低頻的訊號，所以也會經由濾波器來過濾高頻的雜訊，最後將得到的訊號透過微分的方法求得掃視的速度及眨眼的頻率還有開閉眼的時間長短，來推斷駕駛的狀態。

而腦電圖主要是透過電極來測量大腦皮質層的電位變化，可以分為入侵式跟非入侵式，並且腦波通常會用頻率的大小來區分成0-4hz的delta波，4-8hz的theta波，8-13hz的alpha波，13-30hz的beta波，30-100hz的gamma波等等。相比於大部分腦點圖的疲勞駕駛偵測模型，使用support vector machine based posterior probabilistic model，可以不只是用一刀切的方式把是否疲勞駕駛單純的轉化為離散數值，而是根據不同的疲累程度所給予連續區間中的某個值，來定義出駕駛的疲勞程度，如此一來經過提取後的腦電圖特徵及此模型可用來估計駕駛疲勞的程度。最後單獨使用腦電圖做疲勞駕駛偵測主要有三種方式：基於FFT的頻譜分析、基於小波的頻譜分析、基於高階統計的分析，而由於第一種的所需資源最少故最常被使用擷取EEG的特徵。

另有一種基於心跳的方法，是把ECG(心電圖)中的波峰R波之間的時間間隔的記為RRI，這個間隔的波動的程度就是HRV，這種波動是由自律神經系統(ANS)控制的，ANS發送刺激或是放鬆的信號給身體，HRV就是識別ANS的一種方法。

使用HRV測量疲勞駕駛的方法[5]使用了從RRI進行時域分析及頻域分析得出共8個feature，因疲勞駕駛檢測可以視為一種異常檢查問題，MSPC是一種製造業中常用的異常檢測方法，使用MSPC來檢測HRV的異常(也就是昏睡狀態)，對清醒資料模型進行預處理後訂定出control limits後，當駕駛上路會持續計算是否為超過control limits就會判斷進入昏睡狀態，不過，實驗中平均1小時會產生1.7次的false positive，但是因次數不多且false negative比false positive來的嚴重多，可以接受這樣的結果。

從以上得知我們survey了許多種方式來監測駕駛，除了能準確監測駕駛的行為外，如何提醒駕駛也是我們有去調查的問題。目前多種ADAS通過視覺、音頻和振動方式，向駕駛員提供大量安全警報。其中，語音警報被認為是最有效的人類警報形式。研究表明基於語音的駕駛員警報系統已納入大多數用於內部和外部 ADAS 功能的駕駛員警報系統，例如：Kalaivani 等人提供的語音警報生成系統。Sinha，對福特豐田警報系統的研究。J.Källhammer等人提出的針對行人的相對駕駛員警報系統。等等，都會通過檢測車輛、行為、EGC信號、或心理等不同特徵，進行語音回饋。

但在文獻中沒有觀察到有效提高駕駛員對語音警報感知的重要工作。語音警報的靜態特性使駕駛員在情緒異常或駕駛注意力水準下降時瞭解警報的重要性方面效果變差。因此，需要一種專用系統來使語音警報在駕駛員的所有精神狀況下都有效且成功。

4. Formulation

a.MSPC used in HRV-based preparation[5]

i.Formulation of MSPC

Normal data matrix 可進行SVD分解:

$$\begin{aligned} X &= U \Sigma V^T \\ &= [U_R \quad U_0] \begin{bmatrix} \Sigma_R & 0 \\ 0 & \Sigma_0 \end{bmatrix} [V_R \quad V_0]^T \end{aligned}$$

並透過主成分分析(PCA)對X進行矩陣分解:

$$X = T_R V_R^T + E$$

其中 $T_R \equiv U\Sigma$ 為score matrix, $E \in N \times M$ 實矩陣, 為error matrix, V_R 的column生成表示變量之間的相關性的子空間 Π , 其中 $R(\leq M)$ 是PCA模型中保留的主成分的數量。

T^2 統計量用於監測 Π 的異常情況:

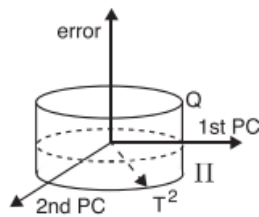
$$T^2 = \sum_{r=1}^R \frac{t_r^2}{\sigma_{t_r}^2} \\ = x^T V_R \Sigma_R^{-2} V_R^T x$$

其中 σ_{t_r} 表示第 r 個score t_r 的標準差, x 是一個新測量的樣本。 T^2 定義為由score標準差的距離標準化。

Q 為樣本與 Π 之間的平方距離:

$$Q = \sum_{m=1}^M e_m^2 = \sum_{m=1}^M (x_m - \hat{x}_m)^2 \\ = x^T (I - V_R V_R^T) x.$$

ii.MSPC model



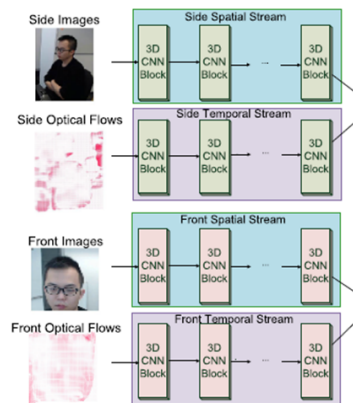
上圖為MSPC中由 T^2 和 Q 定義的normal operating condition(NOC)模型, NOC可以被視為一個圓柱體, T^2 和 Q 統計量的control limits分別對應於其直徑和高度, MSPC會在 T^2 或 Q 超過預定的control limits時檢測到異常情況。因此, 必須仔細確定 T^2 和 Q 。

5. Proposed approach

a.Model ARCHITECTURE

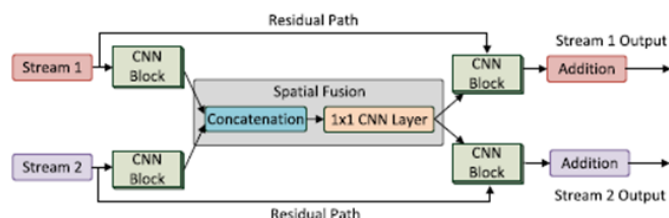
Model使用了四種不同資料作為input, 分別為side video stream, side optical flow, front video stream and front optical flow。在video stream 中每個frame以4個特徵表示, 分別為

time,height,width and RGB channel。在optical flow中每個frame以五個特徵表示, 分別為time,height,width,vertical components,horizontal components。

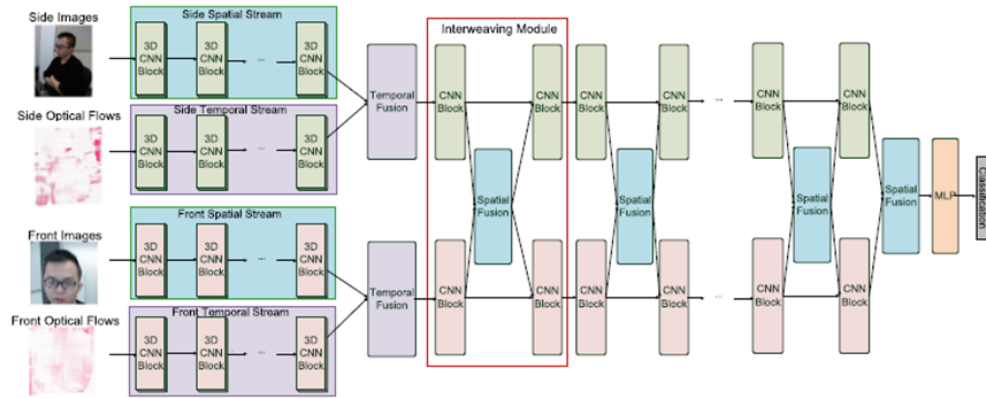


如上圖所示。每種原始資料會經過7層3D CNN blocks, 再經過Batch Normalization layer, 再經過SELU activation layer。四種萃取後的資料分別稱為side spatial stream,side temporal stream,front spatial stream,front temporal stream。

四種stream會分別依照front data and side data進行temporal fusion, fusion根據time axis 進行concatenation operation。在經過temporal fusion後會將front fusion stream and side fusion stream送入Interweaving modules。

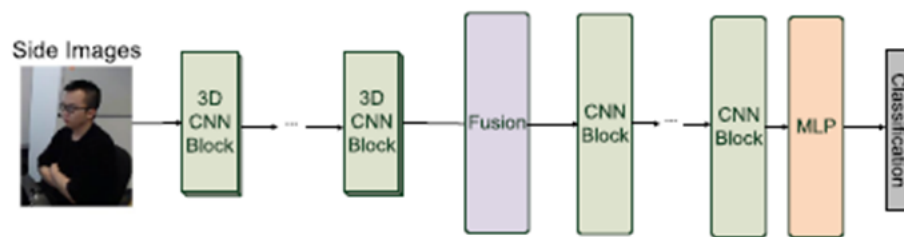


如上圖所示。上圖為Interweaving module structure, 兩個fusion stream先經過一CNN Block使兩個stream能夠進行spatial fusion。在spatial fusion中會將兩個input進行concatenation operation再經過一層1*1的CNN Layer, 這個CNN Layer能夠提升模型的robustness。在經過spatial fusion後會再將data平行地送入兩個CNN Blocks, 同時將兩個fusion stream分別送入後方作為Residual Path, 讓CNN的output與fusion stream相加。在InterCNN中模型使用了25個Interweaving module。

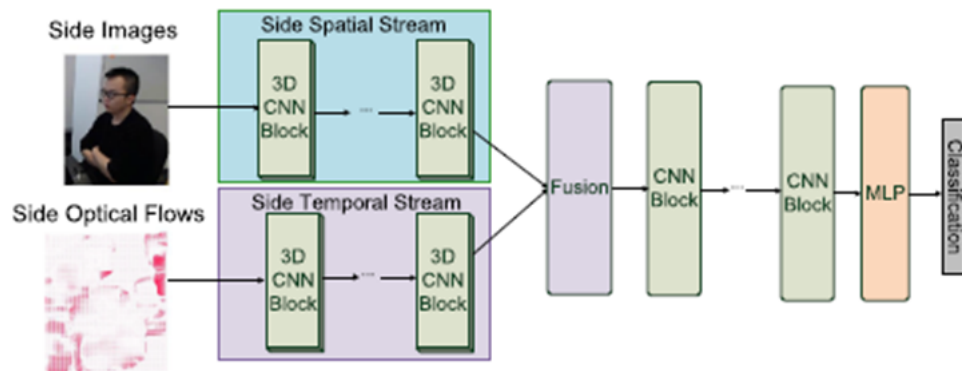


(a) The architecture of the Interwoven CNNs (InterCNNs).

如上圖所示。上圖為InterCNN的structure, 其中CNN Block是影響模型的關鍵因素, 在之後的Exerimental Results中會比較不同model structure及不同CNN Block下的準確率。



(b) The architecture of a plain CNN.



(c) The architecture of a two-stream CNN (TS-CNN).

在本篇論文中會透過三種不同model strucutre進行實驗, 分別為plain CNN,two-stream CNN,InterCNN。上圖為另外兩種model的structure。在plain CNN中會使用side video stream作為input。在two-stream CNN中會使用side video stream及side optical flow作為input。

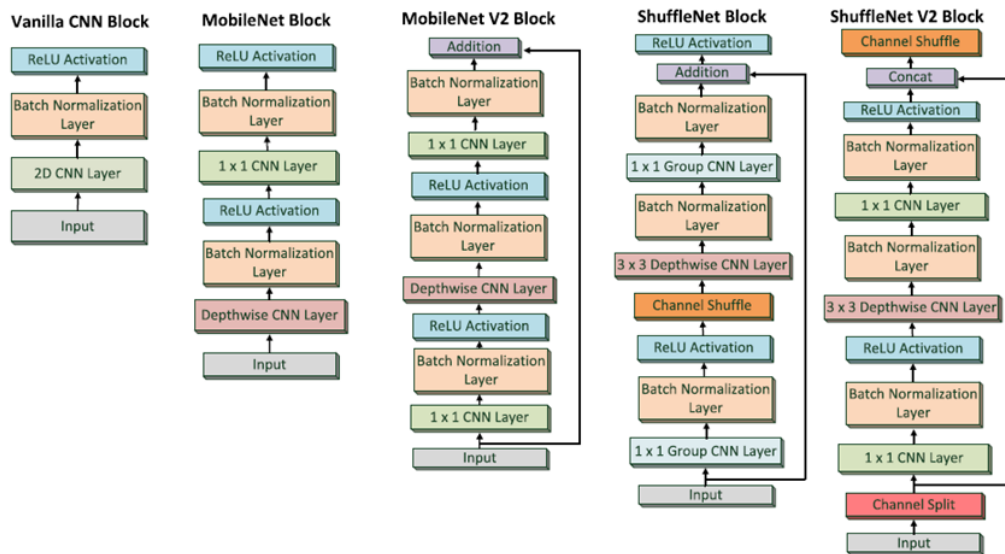
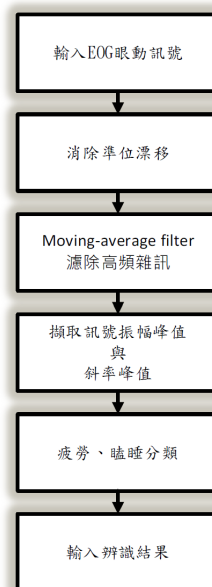


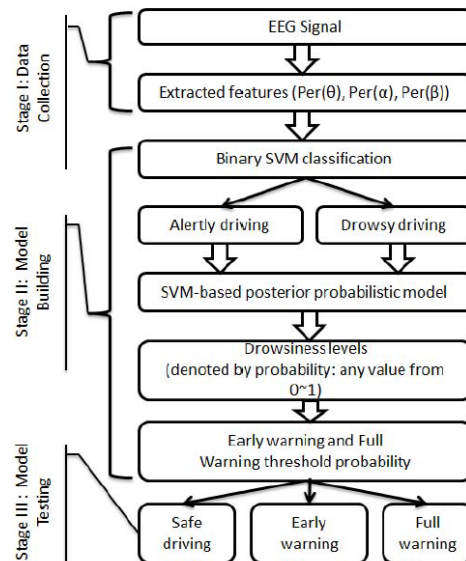
FIGURE 7. The different CNN blocks employed in this study.

在本篇論文中會透過五種不同的CNN Block進行實驗，分別為vanillaCNN,MobileNet,MobileNet V2,shuffleNet,ShuffleNet V2。

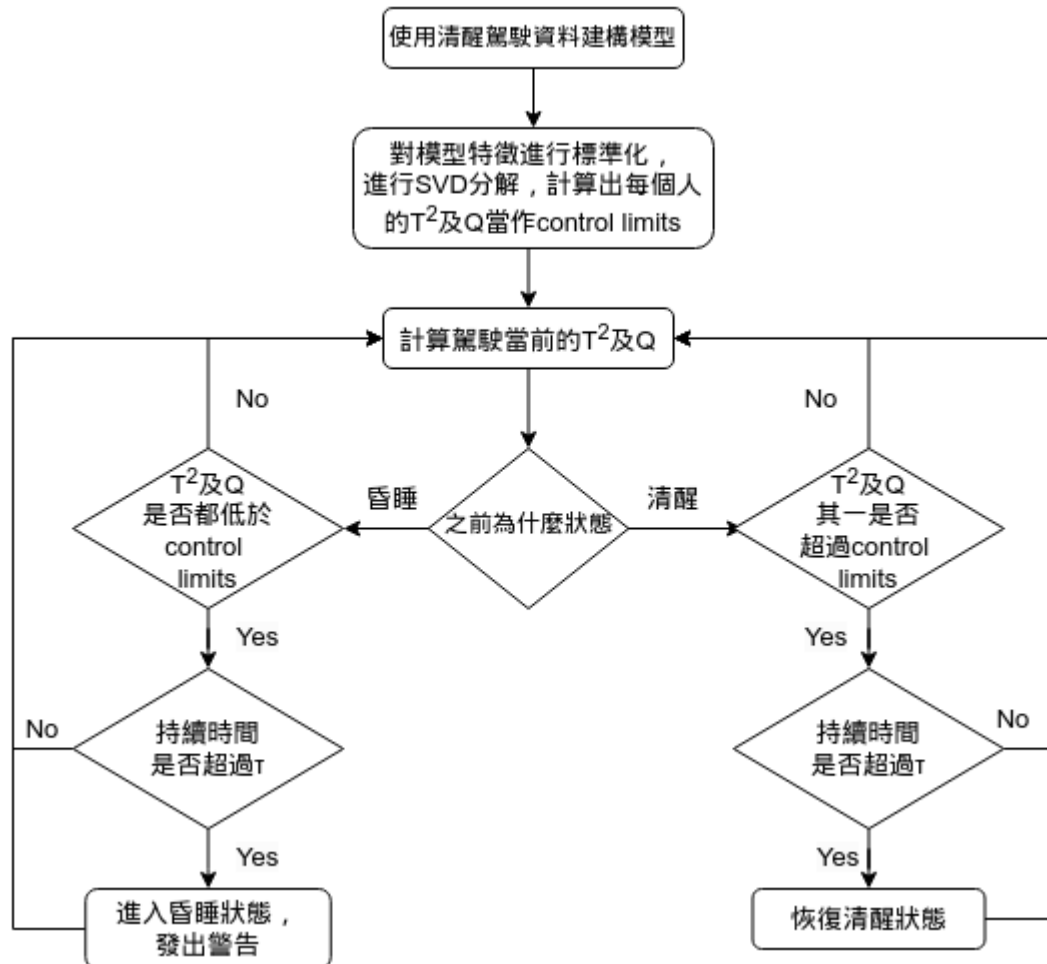
b.Flowchart of EOG system[2]



c.Flowchart of EEG system[1]



d.HRV-based Drowsy Driving Detection



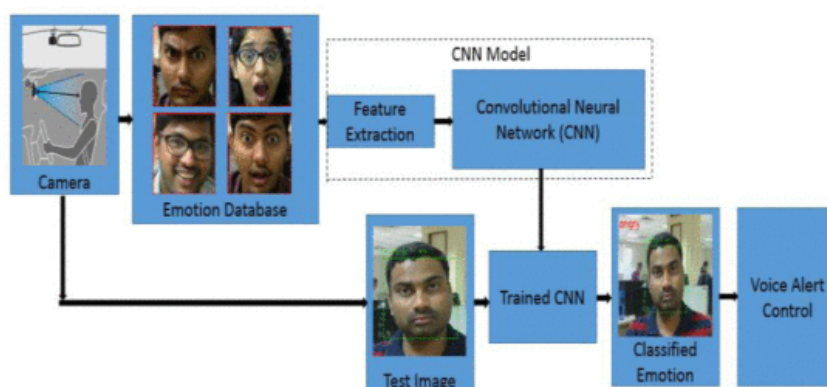
使用HRV來偵測疲勞駕駛時會先用清醒駕駛資料提取8個feature建構模型，再對其進行預處理(標準化)，之後進行SVD分解得出統計量 T^2 and Q 為control limits，因HRV的變化有很大的

個體差異性，會隨著年齡的成長而改變，所以control limits對每個人來說也會是不同的。

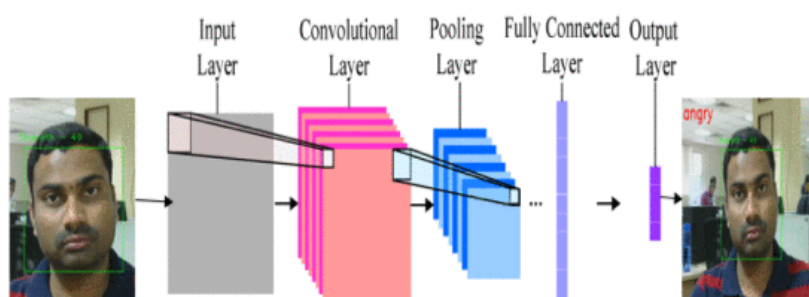
正式上路時，要從身上穿戴的感測器測出的RRI資料計算出 T^2 and Q統計量，因為怕false positive 頻繁發生，我們會訂定一個期間 τ ，若之前保持著清醒狀態，假設測量的 T^2 or Q值超過一開始訂的control limit的持續時間超過 τ ，會認定是昏睡狀態，反之，測量的 T^2 and Q值都低於control limits的持續時間超過 τ ，才會認定駕駛從昏睡甦醒。[5]

e. Emotionally Adaptive Driver Voice Alert System

情緒自適應語音警報系統根據駕駛員的情緒調整警報方式，有效增強警報成功率。



如上圖所示，在ERS 之前，來自前置攝像頭的傳入感測器圖像使用 Open-CV 被處理傳入並使用 Haar-Cascade 分類器檢測人臉並將檢測到的人臉圖像輸入 ERS。檢測到的面部被輸入經過訓練的 CNN，進行情緒分類。最終提供語音警報回饋。其中，系統通過將揚聲器與作為硬體系統佈置的 Raspberry Pi 介面來評估所提出系統的性能。ERS 使用可用的情緒分類標準數據集面部情緒識別 (FER) 2013訓練CNN模型一共將情緒分為：快樂、悲傷、憤怒、恐懼、驚訝、厭惡和中性，七類。人臉識別系統流程圖如下圖所示。



6. Experimental results

a.model performance

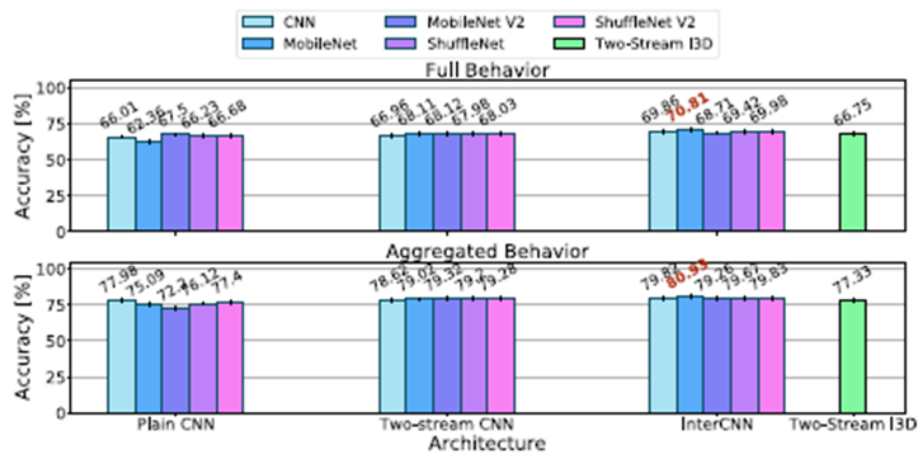


FIGURE 8. Prediction accuracy in the case of classification among 9 different driver behaviors (top) and aggregate tasks (bottom), for all the neural network architectures considered.

論文中實驗了三種不同的模型架構及五種不同的CNN Block。上圖為實驗的結果。綠色為two stream inflated 3D CNN approach, 準確率為66.75%及77.33%, 論文將其視為baseline approach。在上圖可以看出two-stream 及InterCNN都比Plain CNN更好。在Full Behavior中準確率最高的是InterCNN with MobileNet blocks, 準確率為70.81%。在Aggregated Behavior中準確率最高的是InterCNN with MobileNet blocks, 準確率為80.93%。

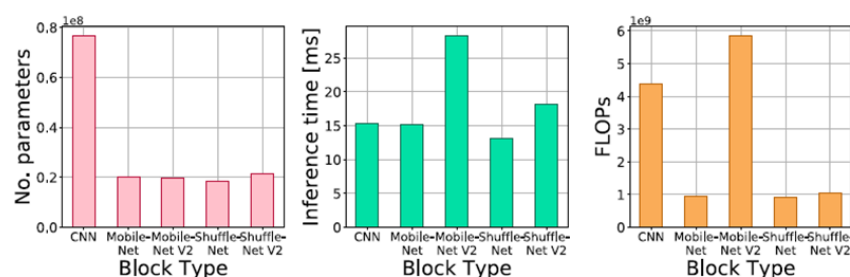


FIGURE 9. Comparison of number of parameters (left), inference time (middle) and FLOPS (right) on InterCNNs with different CNN blocks.

上圖為在InterCNN 下不同block的三種指標。指標分別為 Number of parameters, Inference time, FLOPS。No. parameters 相對於 model size, 會影響 memory and storage on the in-vehicle system。Inference time 相對於 real-time

constraint, 會影響模型是否能應用在車用系統上。FLOPS相關於模型的複雜度, 較高的複雜度通常也代表著越長的running time。Mobile-Net及Shuffle-Net能夠滿足15ms的real-time constraint, No. parameters及FLOPS也很小, 因此被認為是能夠應用在real-world的模型。

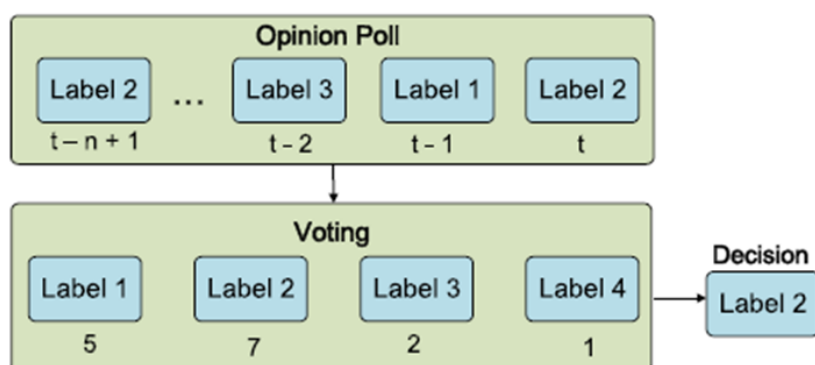


FIGURE 10. Illustration of the temporal voting (TV) scheme.

使用多種input使模型提升了robustness及準確率, 但同時也產生了一些問題, 如何確保兩台camera送來的影像是同步的是最大的問題。論文中提出了temporal voting (TV) scheme迴避了這個問題。上圖為TV scheme的示意圖。TV scheme是透過前n個instance的預測結果輔助預測當下的行為, 透過投票方式輔助決定此時的Label。

TABLE 2. Inference accuracy gained with different CNN blocks over full behaviors, before\after applying the TV scheme.

Block	Plain CNN	TS-CNN	InterCNN
CNN	3.40%	1.23%	1.67%
MobileNet	1.03%	2.45%	3.21%
MobileNet V2	1.98%	2.46%	2.17%
ShuffleNet	2.02%	2.41%	2.20%
ShuffleNet V2	1.84%	2.35%	2.09%

上方表格是模型在不同structure及不同CNN Block下加入TV scheme對full behaviors的結果。在上方表格可以發現TV scheme對每一種組合的準確率都有提升, 平均提升了約2.17%。

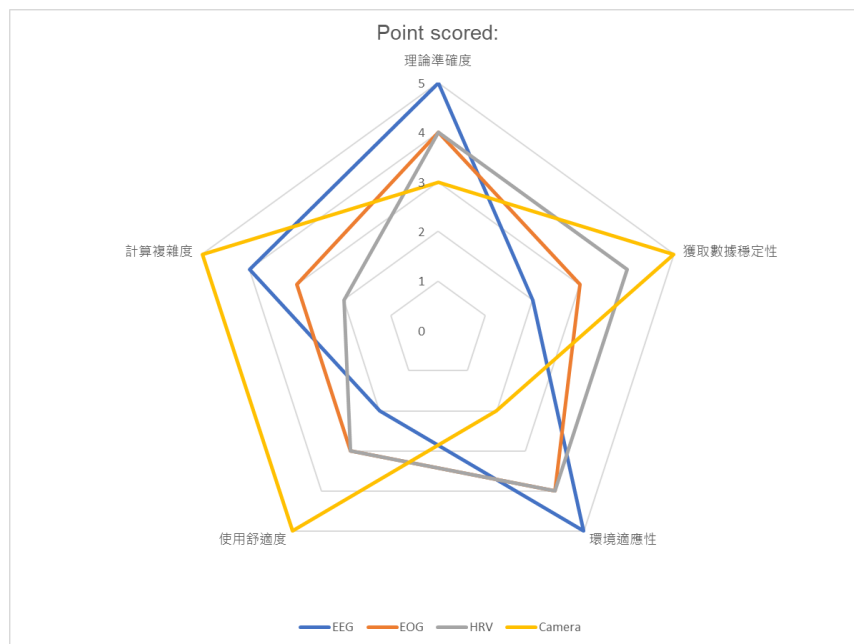
TABLE 3. Inference accuracy gained with different CNN blocks over aggregated behaviors before\after applying the TV scheme.

Block	Plain CNN	TS-CNN	InterCNN
CNN block	1.65%	1.83%	1.90%
MobileNet	2.52%	2.11%	0.98%
MobileNet V2	1.88%	1.55%	1.63%
ShuffleNet	2.23%	2.31%	2.16%
ShuffleNet V2	1.97%	2.19%	2.00%

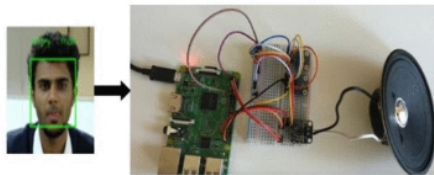
上方表格是模型在不同structure及不同CNN Block下加入TV scheme對aggregated behaviors的結果。在上方表格可以發現TV scheme對每一種組合的準確率都有提升，平均提升了約1.78%。

實驗結果顯示InterCNN+MobileNet能夠達到最高的準確率，同時在memory and storage、Inference time、FLOPS的表現也都是最好的。因此InterCNN+MobileNet被認為是一個能夠被輕易使用在目前車用系統上的solution。

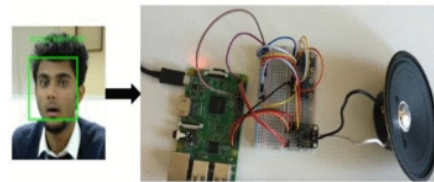
b.comparison of driver monitor method



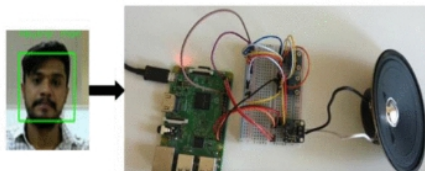
c.Emotionally Adaptive Driver Voice Alert System



Voice Control for Facial 'Angry' Emotion



Voice Control for Facial 'Surprise' Emotion



Voice Control for Facial 'Neutral' Emotion

情緒自適應語音警報系統會持續分析駕駛員的情緒，並在固定的時間內對情緒數據進行分析。基於情緒識別系統結果，系統會相應地生成語音警報。

例如警報控制系統在悲傷、憤怒、厭惡和恐懼等負面情緒狀態下產生強烈的語音警報，在快樂、中性和驚訝等積極情緒狀態下產生適度的語音警報。擬議的語音警報控制系統適應駕駛員情緒的變化，並且由於這種性質，無論駕駛員的情緒如何，它都能夠有效地將警報資訊傳達給駕駛員。

7. Conclusions

駕駛監測系統是會關係到生命安全的重要系統，在經過 survey 過不同的方法監測疲勞駕駛後，我們發現每種方式各有利弊，可以根據自己常遇到的情況來選擇使用何種方法，且透過在車上或是駕駛身上穿戴額外的感測器來收集數據。我們認為最理想是每一種都使用這樣一來各種情形都會考慮到，不但能達成行車安全最大化這樣的目的，且不致對車輛或駕駛有太大的負擔。

未來自駕車若能普及關於駕駛的不良行為或是疲勞駕駛可能不是最需要擔心的，但我們認為若遇到緊急狀況或駕駛身體出現一些狀況時，又或者要根據駕駛的行為及表情，車輛能辨別出駕駛需要的服務，駕駛監測系統在其中都扮演著一份重要的角色。提高檢測的準確性與給予個性化的回饋，能夠使 HMI 更加複雜，並使 ADAS 系統通過持續平穩的駕駛活動更有效地實現零事故願景。

8. Acknowledgement

洪郡辰	Camera-based Driving Detection with InterCNN
周柏諺	EOG & EEG based Drowsy Driving Detection
張宏洋	HRV-based Drowsy Driving Detection
何樂為	Emotionally Adaptive Driver Voice Alert System

9. Reference:

1.C. Zhang, R. Li, W. Kim, D. Yoon and P. Patras, "Driver Behavior Recognition via Interwoven Deep Convolutional Neural Nets With Multi-Stream Inputs," in *IEEE Access*, vol. 8, pp. 191138-191151, 2020, doi: 10.1109/ACCESS.2020.3032344.

2.張袁。「基於眼電圖法的疲勞駕駛偵測」。碩士論文，國立中央大學電機工程學系，2017。

3.Wang FWXu QFu RRStudy on the effect of man-machine response mode to relieve driving fatigue based on EEG and EOGSensors201910.3390/s19224883

4.G. Li, B. -L. Lee and W. -Y. Chung, "Smartwatch-Based Wearable EEG System for Driver Drowsiness Detection," in

IEEE Sensors Journal, vol. 15, no. 12, pp. 7169-7180, Dec. 2015, doi: 10.1109/JSEN.2015.2473679.

5.K. Fujiwara *et al.*, "Heart Rate Variability-Based Driver Drowsiness Detection and Its Validation With EEG," in *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1769-1778, June 2019, doi: 10.1109/TBME.2018.2879346.

6.S. M. Sarala, D. H. Sharath Yadav and A. Ansari, "Emotionally Adaptive Driver Voice Alert System for Advanced Driver Assistance System (ADAS) Applications," 2018 International Conference on Smart Systems and Inventive Technology (ICSSIT), 2018, pp. 509-512, doi: 10.1109/ICSSIT.2018.8748541.