

基於ShuffleNet之道路事件聲音識別輕量化邊緣智慧系統

指導教授: 陳聿廣

學生姓名: 戴仕庭、鄭凱方、吳育丞、呂翊銓

摘要：為因應日益增長的交通事故，我們提出道路事件聲音識別系統，以提高既有監視系統的全面性和辨識正確率。本專題使用單一麥克風實時接收時域信號，再運用MFSC轉換成頻譜信號，最後透過ShuffleNet在FPGA上實現即時辨識，這將有助於加速道路救援行動。

一、前言

現今仍存在的難題：

- 道路事件發生頻率急速增加
- 道路事件的通報流程繁瑣
- 監視器存在視覺死角

因此我們提出—基於ShuffleNet之道路事件聲音識別輕量化邊緣智慧系統，達成交通事故即時監測。



Fig.1 交通事故數量

二、應用場域

在實際應用中，我們的系統將持續實時接收周遭的聲音訊號，並進行即時分析。當發生車禍等道路事件時，一旦系統識別出撞擊聲，便會立即啟動緊急通報機制，能顯著提升道路救援的效率。

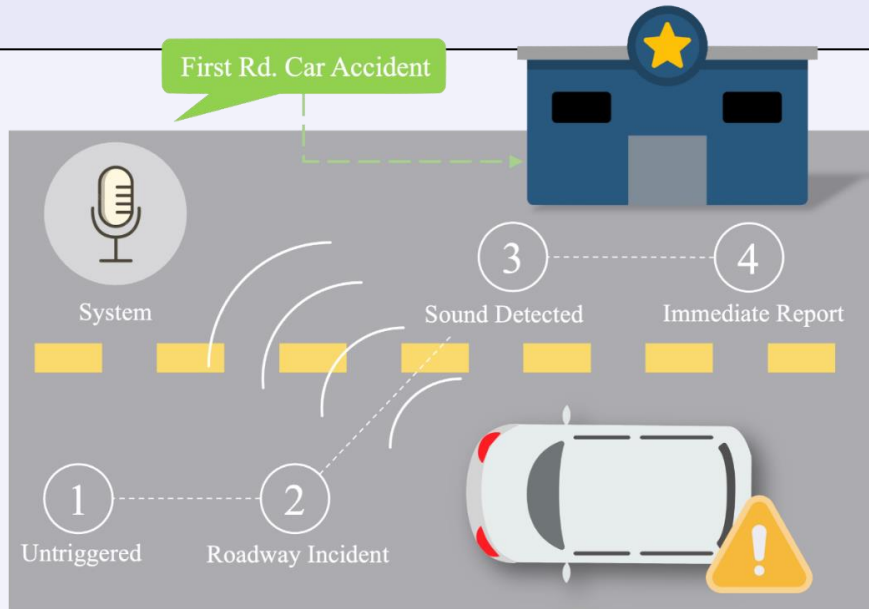


Fig.2 系統概念圖

三、研究方法

本專題之系統為軟硬體整合開發，其流程如下圖 (Fig.3)所示：

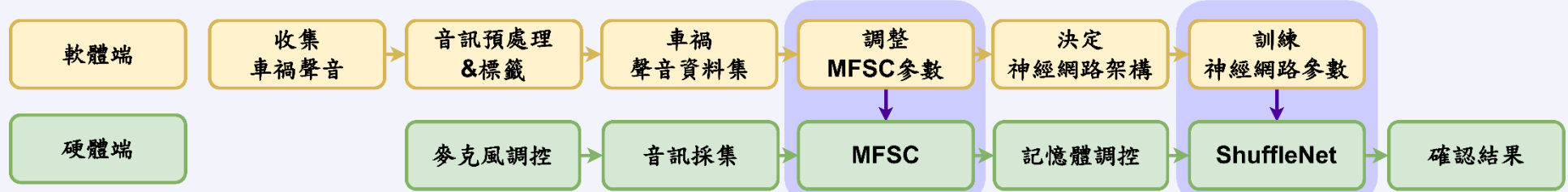


Fig.3 研究流程圖

軟體端：建立、前處理資料集並訓練高準確度的ShuffleNet神經網路。
硬體端：採用PYNQ-Z2型號FPGA，外接AT9901麥克風接收音訊，經由聲音預處理MFSC將時域訊號轉為頻域訊號，最終在記憶體控制與狀態機調控下，將頻譜圖傳入ShuffleNet神經網路進行辨識。

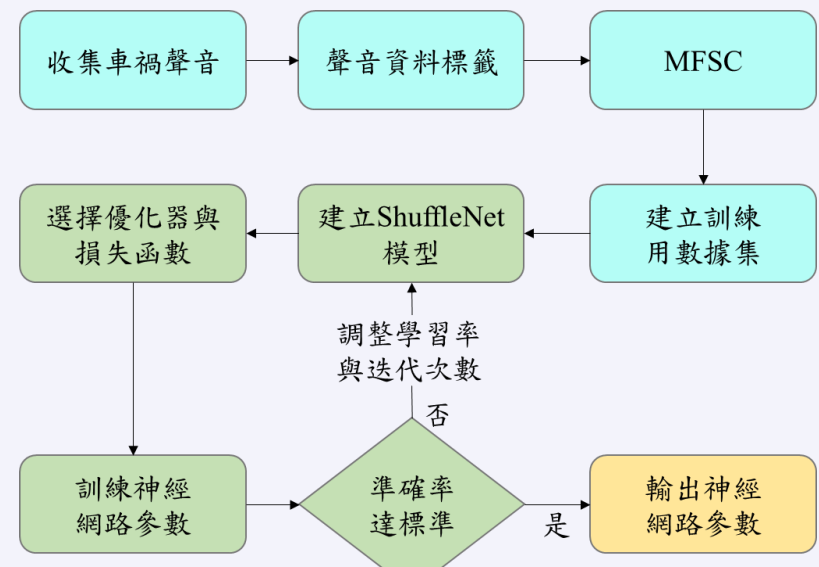


Fig.4 軟體流程圖

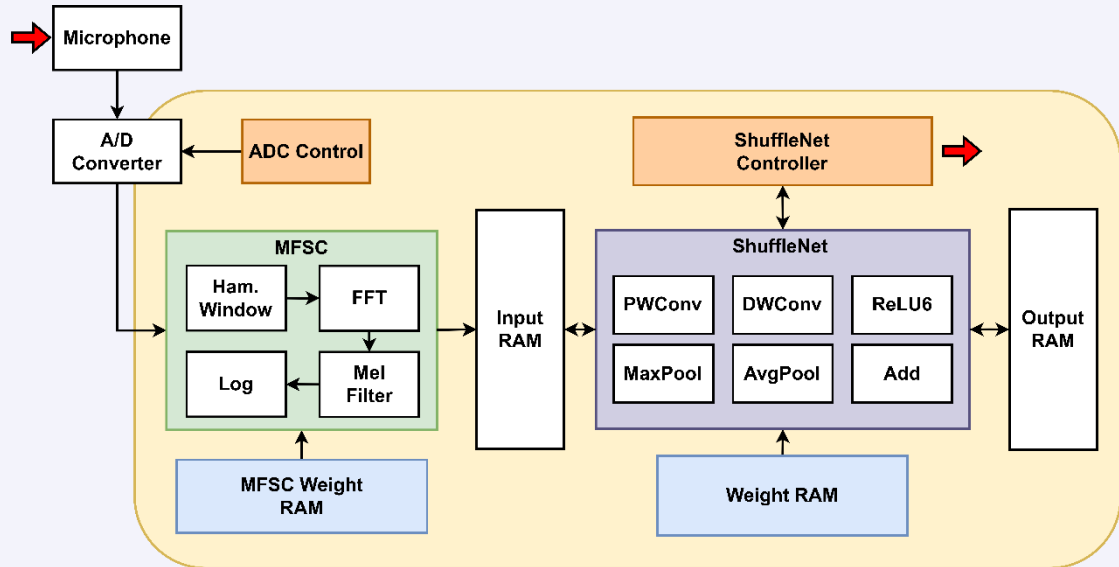


Fig.5 硬體架構圖

以下是MFSC流程圖(Fig.6)，其根據人耳聽覺特性設計，為非線性濾波，有助於去除特定頻率區域內的資料，從而獲得更具代表性的頻譜特徵。

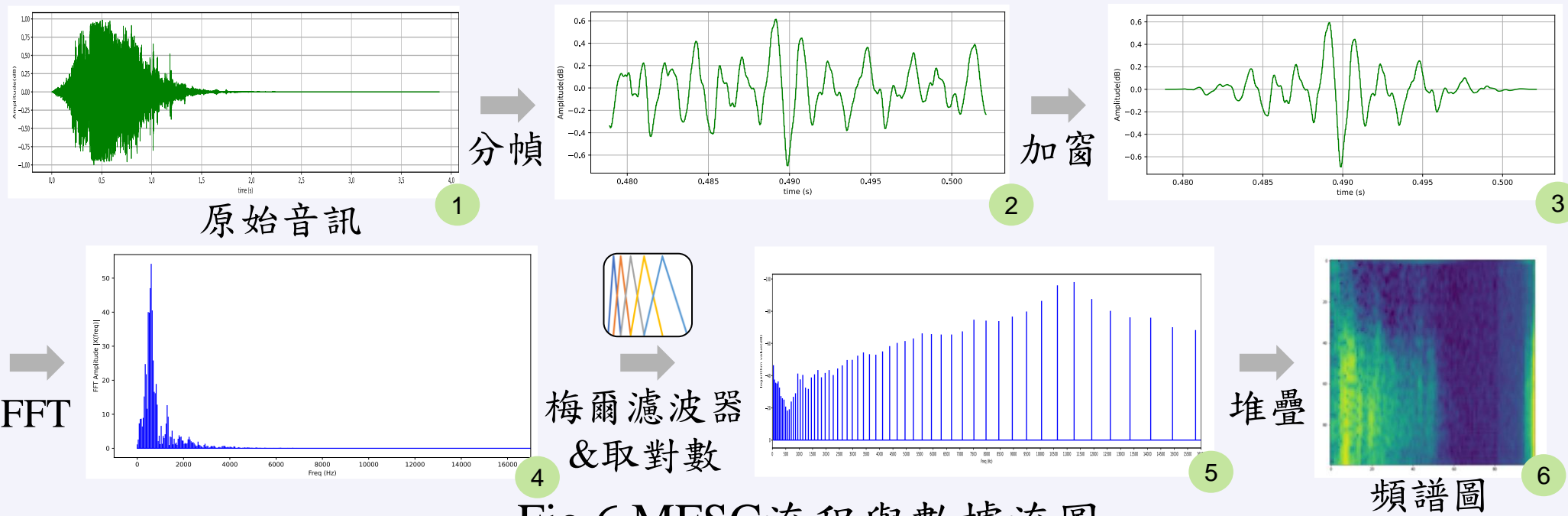


Fig.6 MFSC流程與數據流圖

以下是我們利用Vivado所設計之硬體系統方塊圖(Fig.7):

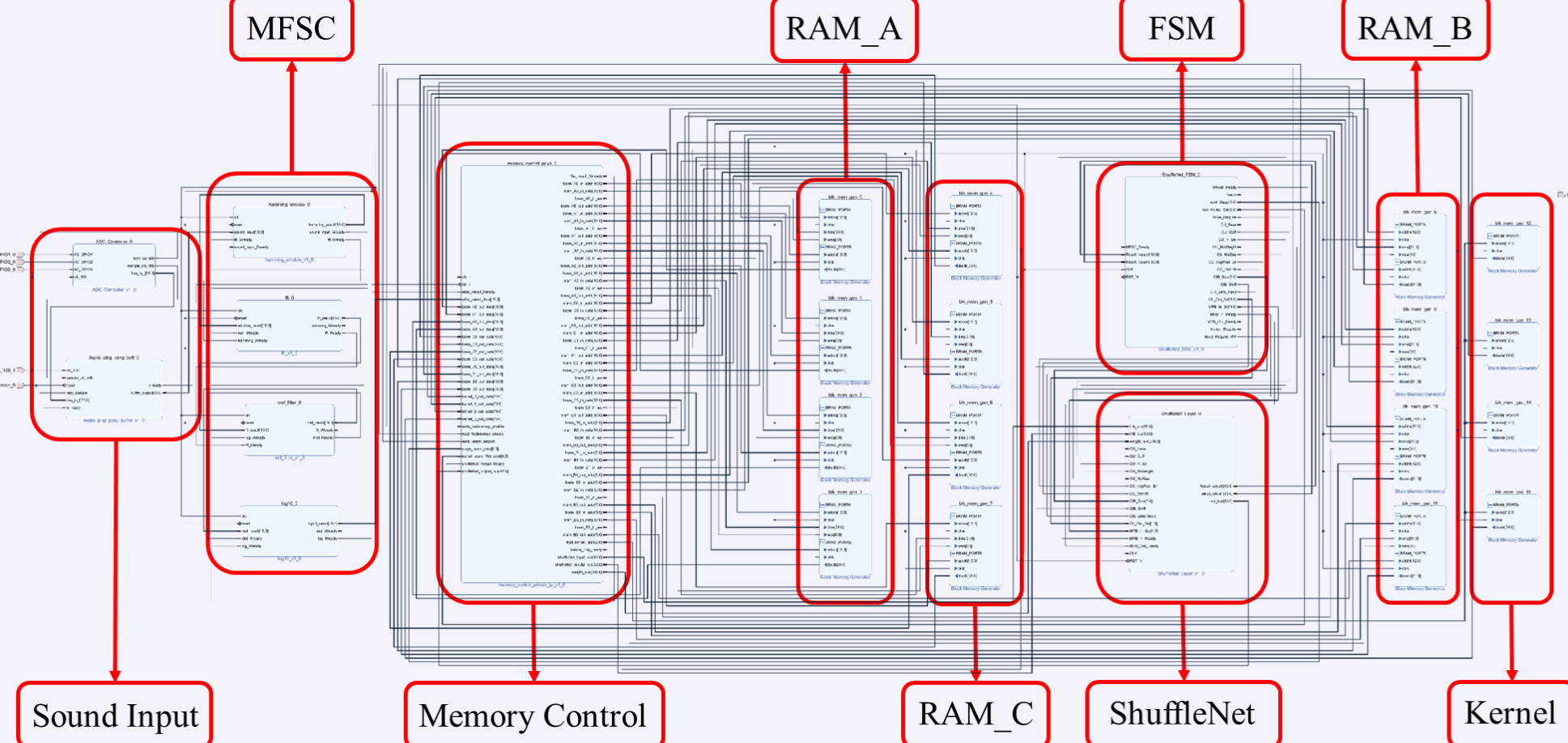


Fig.7 硬體系統方塊圖

四、結果

準確度高達92.7%
Loss為0.378

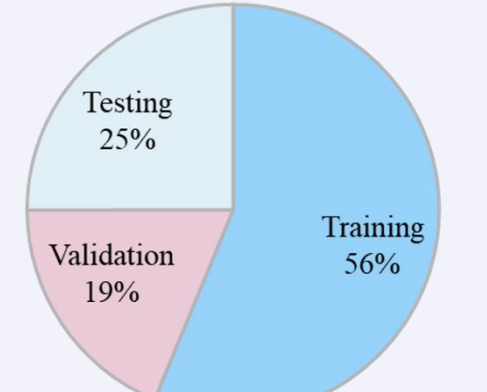


Fig.13 資料集分布圖

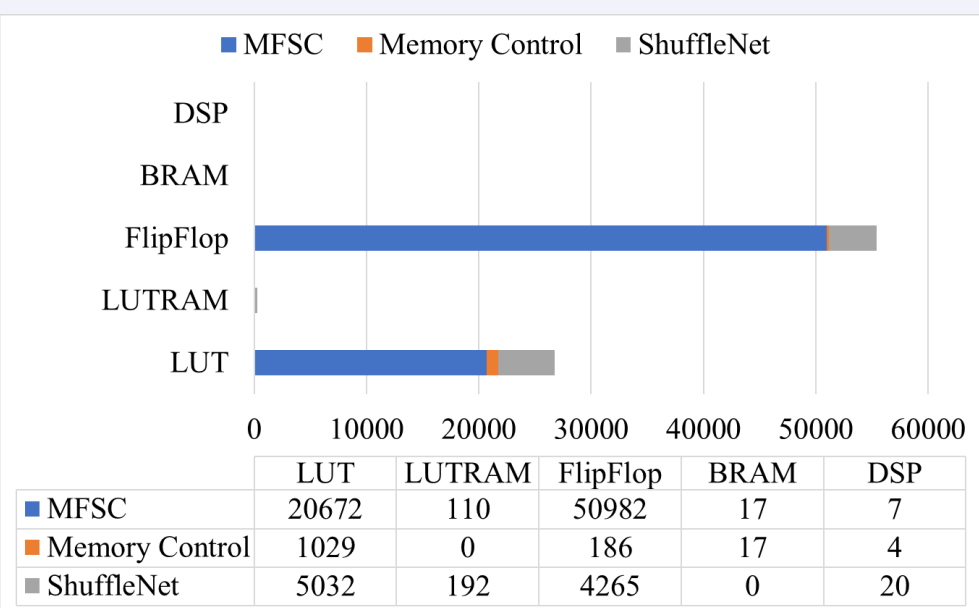


Fig.14 系統各部分資源使用量



成果展示

ShuffleNet是一種輕量化的模型，其採用了深度可分離卷積(包含逐點卷積PW和深度卷積DW)以及通道洗牌技術，能在保持準確度同時以極低參數量和較高運算速度進行辨識。

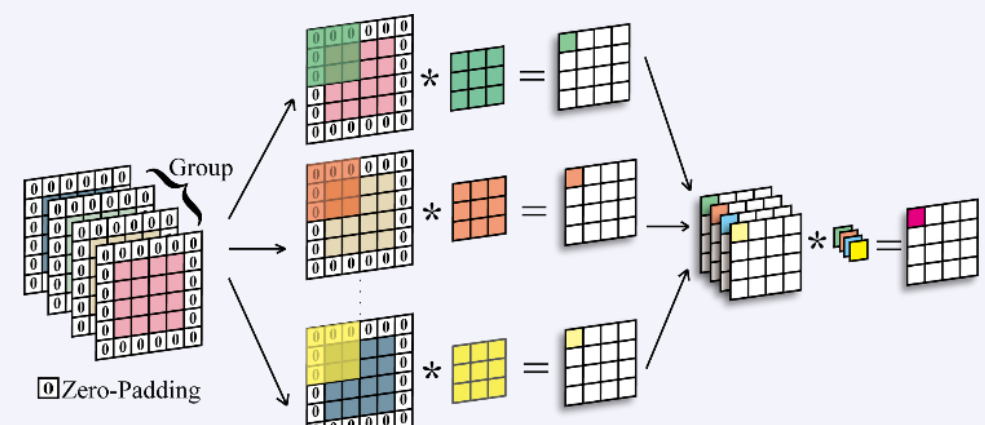


Fig.8 深度可分離卷積

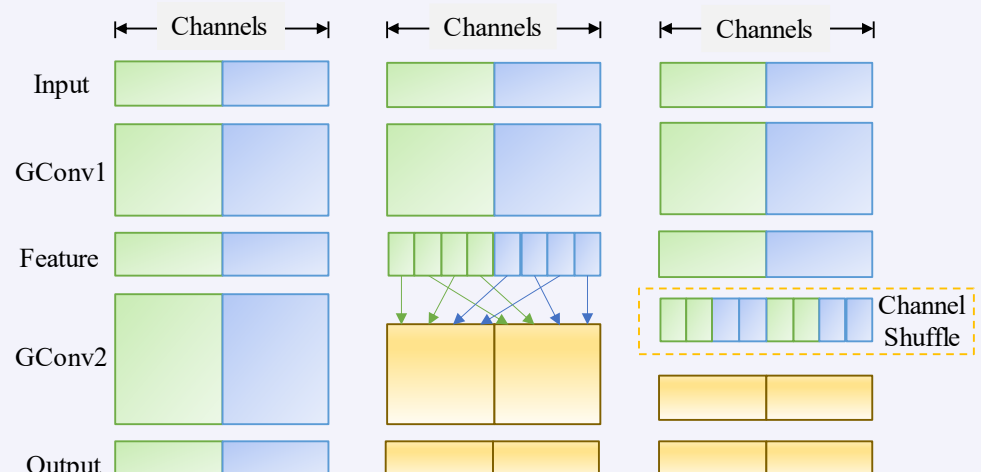


Fig.9 通道洗牌(Channel Shuffle)

如果要在邊緣系統實現即時音訊辨識，必須考慮硬體的資源使用量以及神經網路的運算速度。為此，我們對硬體架構進行了以下優化：

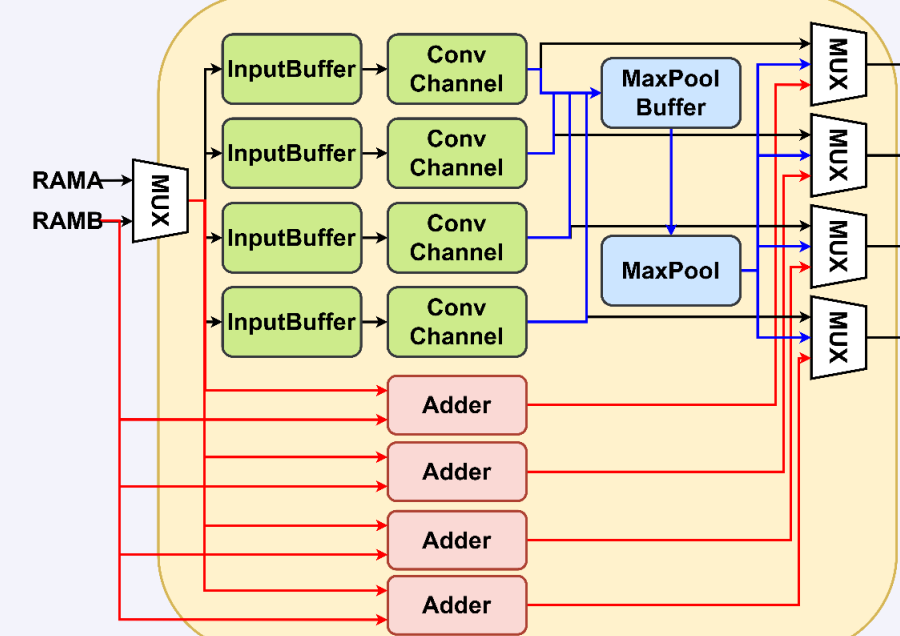


Fig.10 ShuffleNet計算層

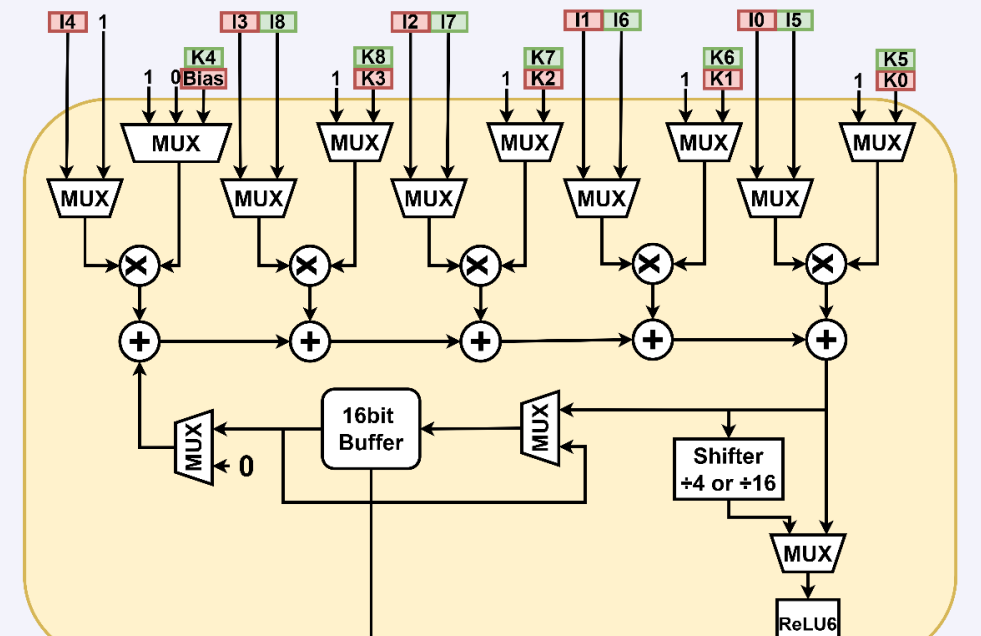


Fig.11 卷積通道

- 平行運算 Parallelism & 運算資料共享(Fig.10)
為滿足hard real-time需求，我們採用四組平行運算單元進行加速，並實現輸入與卷積核共享，有效降低30%以上的記憶體使用量。
- 流水線Pipeline & 乒乓技術Ping-Pong (Fig.7,10)
為確保數據處理的連續性和效率，我們在Pipeline中加入Ping-Pong Buffer，以克服記憶體的存取衝突。同時，InputBuffer則為特化之Taped delay line設計，針對深度卷積DW的特徵圖存取進行優化。
- 運算步驟合併 & 硬體複用 (Fig.10,11,12)
為實現系統輕量化，我們將Global Conv和MaxPool步驟合併，成功降低20%以上的記憶體使用量。再者，我們設計了可高度複用之卷積通道，支持ShuffleNet中的多種運算，以適應邊緣裝置需求。

Layer		Stride	Output Size	Output Channel
Input		-	64x64	1
3x3 Global Conv		2	32x32	16
MaxPool		2	16x16	16
Stage1	Block1	2	8x8	32
	Block2	1	8x8	32
Stage2	Block1	2	4x4	64
	Block2	1	4x4	64
	Block2	1	4x4	64
	Block2	1	4x4	64
Global AvgPool		-	1x1	64
Fully Connected		-	2	1

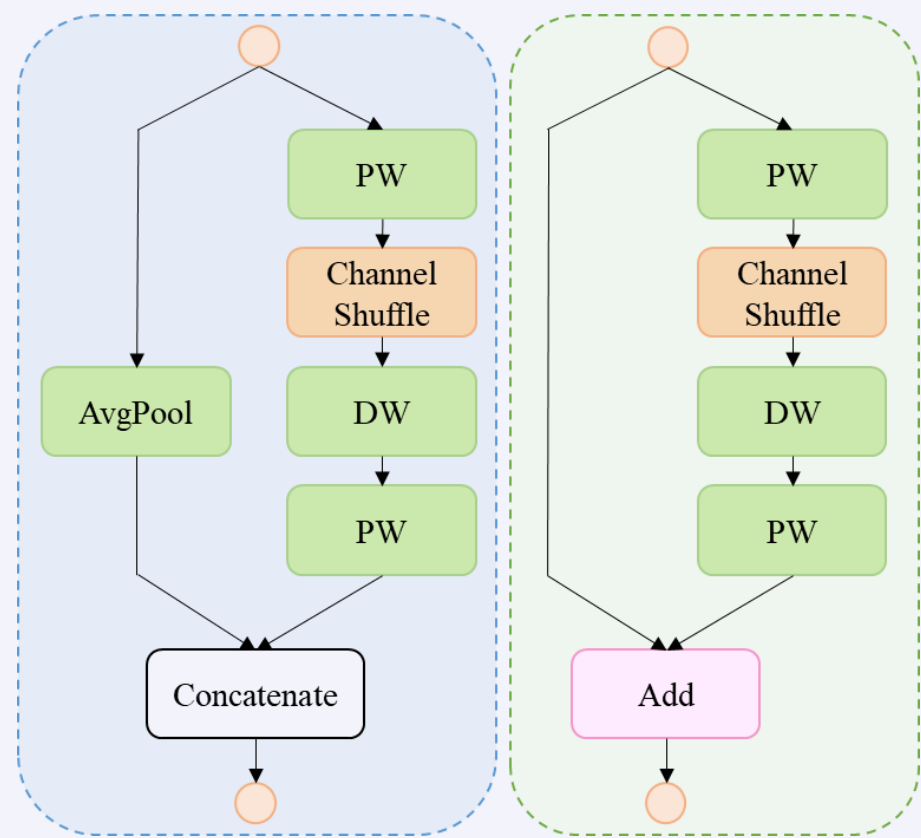


Fig.12 ShuffleNet神經網絡架構、Block1(左)、Block2(右)

五、討論與結論

本專題融合軟硬體整合開發，先由軟體端訓練出高準確度的ShuffleNet神經網路，再由硬體端透過平行運算、資料共享、運算步驟合併、硬體複用等技術，成功實現輕量化的車禍聲音辨識。未來，我們希望將影像辨識整合，透過聲音與影像的結合，我們能更全面地偵測交通事故，為緊急救援工作贏得關鍵的時間。

六、參考文獻

- X. Zhang, X. Zhou, M. Lin and J. Sun, "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 2018, pp. 6848-6856, doi: 10.1109/CVPR.2018.00716.
- Peng, P., et al., Design of an Efficient CNN-Based Cough Detection System on Lightweight FPGA. IEEE Transactions on Biomedical Circuits and Systems, 2023. 17(1): p. 116-128.