



◆ D24-065 ◆

## 應用於即時影像辨識之高效能 PoolFormer DNN處理器

An Energy-efficient PoolFormer DNN Processor  
for Real-time Image Classification

隊伍名稱 | 來去夏威夷 Let's Go to Hawaii

隊長 | 陳永泰 / 清華大學電機工程研究所

隊員 | 廖泓全 / 清華大學電機工程研究所  
蔡惠芸 / 清華大學半導體研究學院  
張愷峰 / 清華大學半導體研究學院

◆ 指導教授 ◆



黃朝宗 | 清華大學電機工程學系

臺灣大學電子工程博士，現為清華大學電機工程學系副教授。曾服務於聯詠科技，亦曾於麻省理工學院進行博士後研究。曾獲傑出人才基金會年輕學者創新獎、未來科技獎、中國電機工程學會優秀青年電機工程師獎、清華大學傑出教學獎、旺宏金矽獎最佳指導教授獎等獎項。

研究領域

近年來研究以實現高效能、高品質之電腦視覺與計算攝影學應用為主，包含卷積神經網路處理器、立體3D光場顯示器、光場相機等相關研究，是國內極少數能同時發表頂尖論文至計算機架構 (ISCAMICRO)、晶片設計 (ISSCC / VLSIC / ESSCIRC)、及電腦視覺 (CVPR / ICCV / TPAMI) 等三大熱門領域之研究者。

◆ 作品摘要 ◆

近年來，深度學習 (Deep Learning) 技術迅速發展，在各種電腦視覺應用中展示了卓越的性能，如影像辨識 (Image Classification)、物件偵測 (Object Detection) 和影像分割 (Image Segmentation)。這些技術的進步主要得益於強大的計算能力和先進的模型架構。其中，通道多層感知機 (Multi-layer Perceptron, MLP) 在Transformer等先進模型中扮演了重要的角色。例如，基於MLP結構的模型MetaFormer通過池化運算 (Pooling) 實現了高效的空間資訊交換，這在低運算量和小模型尺寸的限制下，顯著提高了模型的推論準確率。

MetaFormer這類模型因其高效的推論特性，有望在嵌入式電子產品中實現性能和準確率的全面提升。然而，這些模型對高頻寬和特殊運算架構的需求，仍然對在一般行動裝置和消費性電子產品中的應用造成了一定的挑戰。

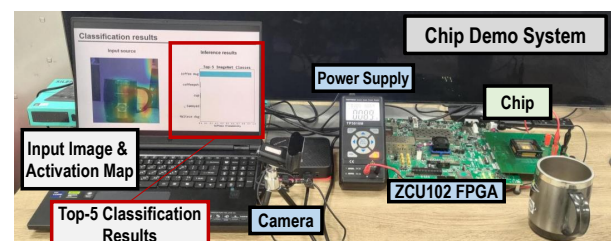
在此背景下，我們提出了一個高效能的即時影像辨識加速器晶片。這個晶片專門針對MLP運算架構進行了硬體最佳化設計，通過改善運算流程，克服了頻寬需求的瓶頸。此外，我們還為特殊運算設計了專門的加速電路，使得這款加速器晶片能夠在嵌入式電子產品上達到更高的準確率。這樣的設計不僅提高了推論的速度，還顯著降低了功耗，使其更適合在電池供電的設備上使用。

為了驗證我們的設計，我們利用Xilinx ZCU 102 SoC開發板進行了即時影像分類的展示。我們的實驗顯示，這款加速器晶片能夠即時顯示辨識結果和注意力集中區域，實現了在邊緣裝置上運行高品質模型推論的願景。

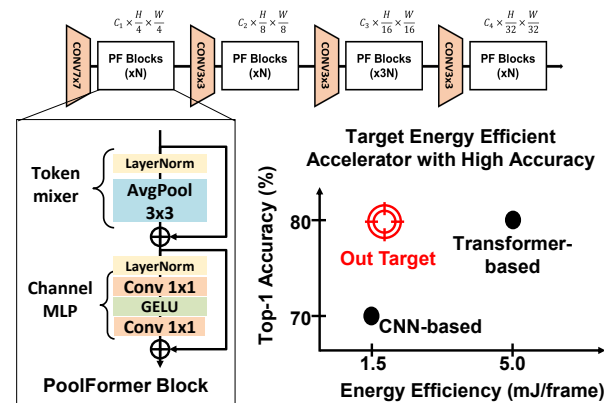
本晶片使用16nm FinFET先進製程下線，支援高精度影像辨識。在實驗量測中，這款晶片以超低功耗實現了超過80%的Top-1 ImageNet辨識正確率。此設計在不犧牲性

能的情況下，顯著降低了功耗，達到了在嵌入式裝置上運行高效深度學習模型的目標。

總結而言，本計畫成功展示了一款高效能且高正確率的影像辨識晶片，能夠在邊緣裝置上實現高品質的辨識功能。我們的研究不僅證明了MLP結構在嵌入式系統中的潛力，還為未來的嵌入式深度學習應用提供了一個有效的解決方案。隨著技術的不斷進步，我們期待這些嵌入式系統能夠在更多的應用場景中發揮作用，推動人工智慧技術的普及與應用。



圖一 即時影像辨識演示系統。



圖二 Poolformer架構與設計目標。

◆ Abstract ◆

In recent years, deep learning has rapidly evolved, showing exceptional performance in various computer vision applications such as image classification, object detection, and image segmentation. Among the notable advancements, the Channel multi-layer-perceptron (MLP) has emerged as a crucial network in transformer-based models. Furthermore, Metaformer facilitates efficient spatial information exchange by leveraging pooling operations, significantly enhancing model inference accuracy under low computational power and compact model size. This improvement has propelled Metaformer to achieve state-of-the-art performance in applications like image classification.

Deep learning models using MLPs have demonstrated great promise for enhancing performance and accuracy in embedded electronic devices. These devices can benefit from the sophisticated capabilities of MLPs, enabling more precise and efficient processing of visual data. However, the implementation of MLPs on general mobile devices and consumer electronics poses significant challenges. These challenges stem from the MLPs' requirements for high bandwidth and specialized computational architecture, which are not typically available on standard consumer devices. Overcoming these obstacles is essential to fully realize the potential of MLPs in practical, everyday applications.

To address these challenges, our project proposes the development of a high-performance real-time image recognition accelerator chip specifically optimized for the MLP's computational architecture. The design of this chip focuses on refining the computational process to overcome bandwidth constraints and implementing hardware tailored for specialized computing operations. By optimizing the chip's architecture, we can significantly enhance the efficiency and speed of MLP-based computations, making them feasible for use in everyday consumer electronics.

Our project also integrates this accelerator chip with the Xilinx ZCU 102 SoC FPGA development board to demonstrate real-time image classification capabilities. The FPGA demonstration system cooperate

with the refined computational process for the MLPs, ensuring that the chip can handle the computational load effectively. This setup allows us to showcase the practical applications of our design by directly displaying recognition results on the screen. The demonstration system serves as a platform for validating the real-time performance of our chip, highlighting its potential in executing high-quality model inference on edge devices.

In summary, our project presents a high-performance, high-accuracy image recognition chip capable of delivering advanced recognition functionality on edge devices. By addressing the key challenges associated with the deployment of MLPs in consumer electronics, we pave the way for broader adoption of deep learning technologies in everyday applications. Our work promises significant improvements in user experience and application performance, bringing the sophisticated capabilities of deep learning models to a wider range of devices. This advancement holds the potential to transform the way visual data is processed and utilized in various consumer electronics, enabling smarter and more efficient devices that can handle complex tasks with ease.

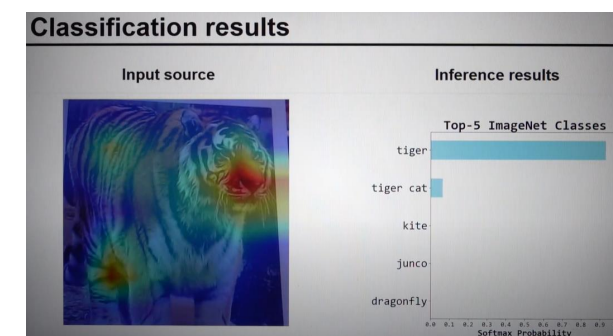


Fig. 3 Classification result.