國立清華大學
NATIONAL TSING HUA UNIVERSITY

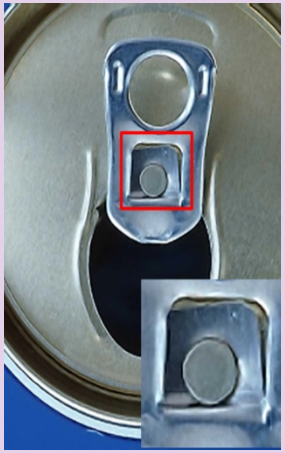
Multimedia and Network System on Chip Lab

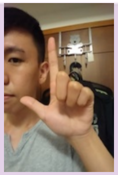
邱瀞德 教授

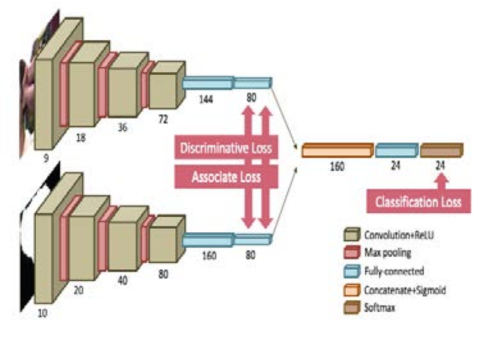
國立清華大學資訊工程學系/通訊所/資應所所長

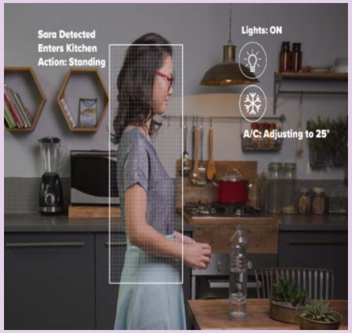
Research Interests:

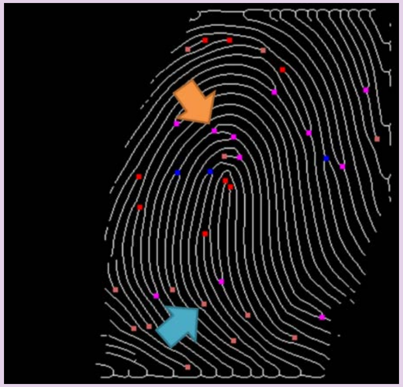
- Pattern Recognition
  - Face/Gesture/Fingerprint recognition
  - Action Recognition/Gait recognition
- Machine Learning
  - 3D Reconstruction/RGB-D
    Object Recognition
  - Depth Prediction
- Image/Video Processing
  - High Dynamic Range
  - Super-resolution
- Hardware Optimization and Implementation

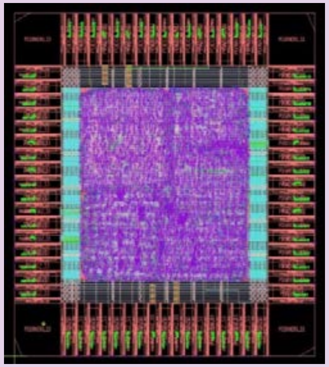


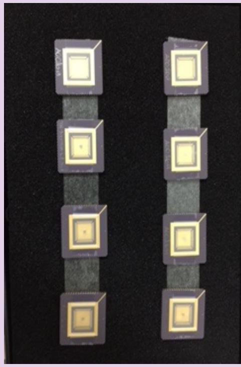










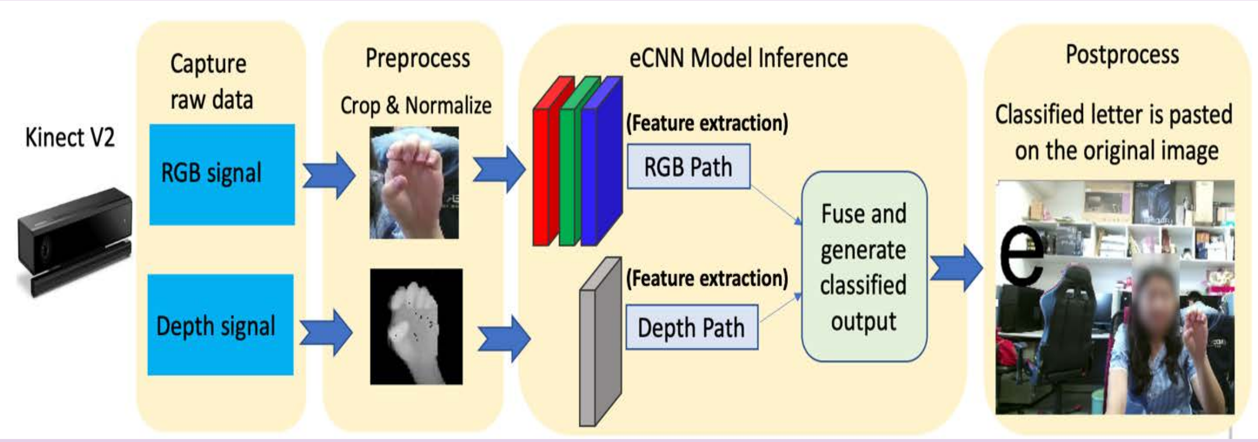


NATIONAL TSING HUA UNIVERSITY

國立清華大學

FAST AND ACCURATE EMBEDDED DCNN FOR RGB-D BASED GESTURE RECOGNITION

- RGB-D based sign language recognition
- Adding depth images increases accuracy around 10%
- Model was designed in hardware-oriented manner for compatible implementation on CNN accelerator
- Our proposed DCNN model outperforms the state-of-the-arts in parameters usage **0.17M** and in **99.79%** accuracy of ASL Finger Spelling dataset
- Fast inference times by RTL simulation and at GTX 1080 are **0.171 ms** and **14 ms**


Kinect V2

Capture raw data

RGB signal

Depth signal

Preprocess

Crop & Normalize

eCNN Model Inference

(Feature extraction)

RGB Path

(Feature extraction)

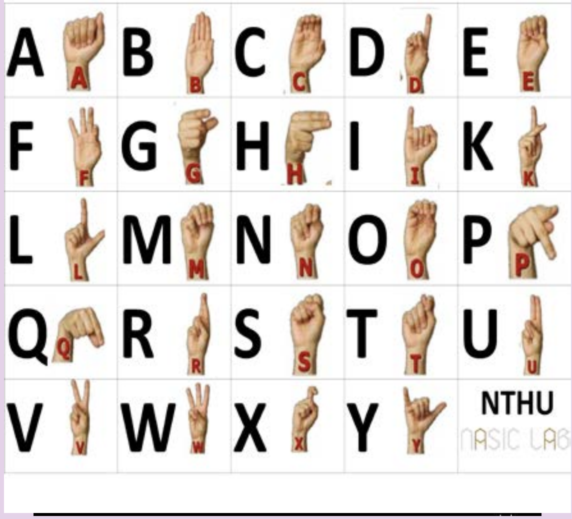
Depth Path

Fuse and generate classified output

Postprocess

Classified letter is pasted on the original image

Fig. 2. Overall sign language recognition system

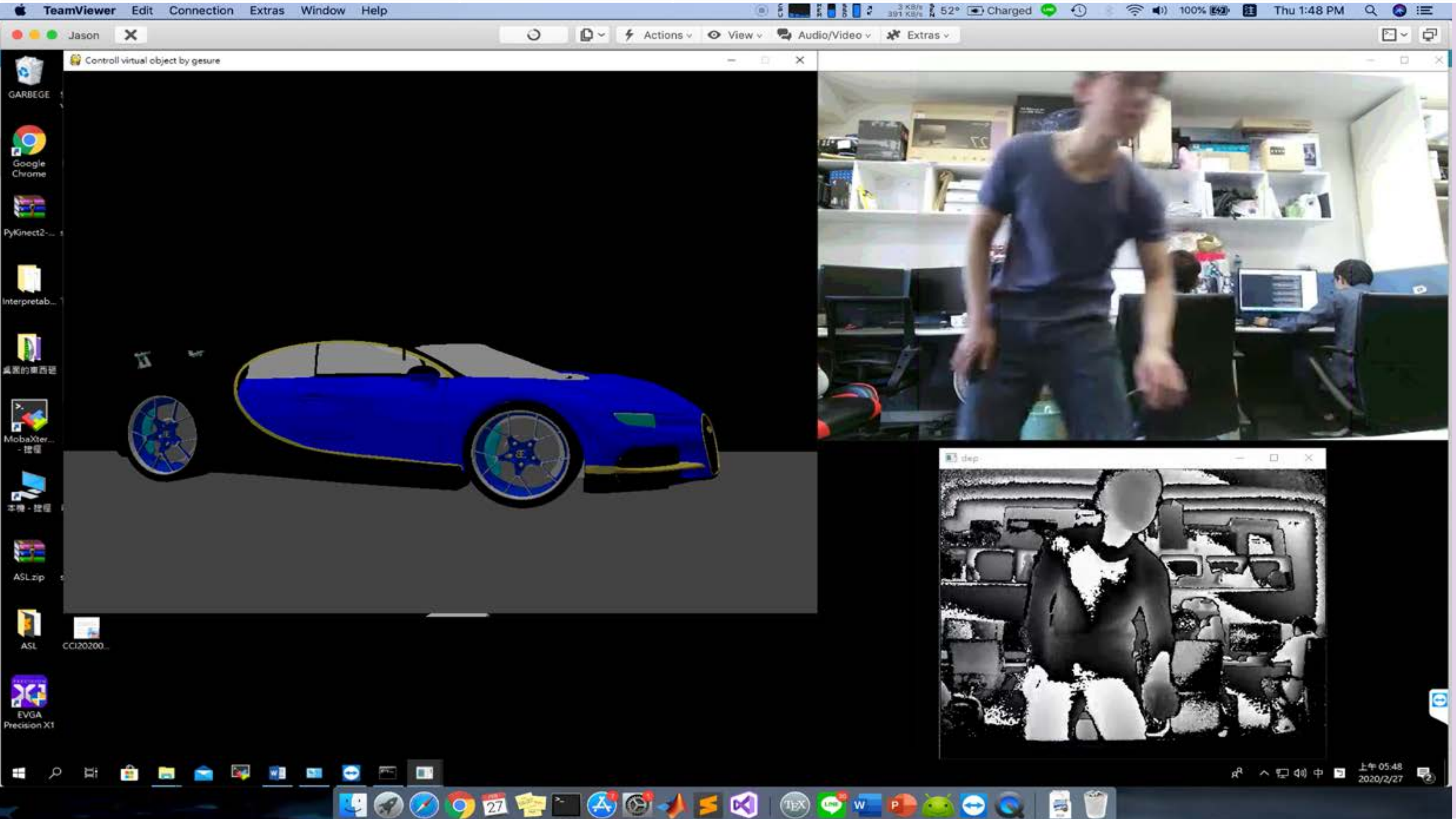


國立清華大學
NATIONAL TSING HUA UNIVERSITY

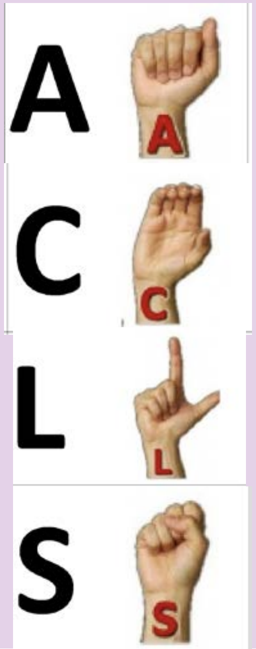
Virtual 3D Object Control with Sign Language Gesture

Model	Pugeault	Rodrguez	Gao	Ma	Li	Ours
Method	Random forest	SVM	CNN	Deep Belief Net	SAE+PCA	ECNN
Accuracy (%)	75	91.26	93.3	96.14	99.10	99.79

Performance comparison



A: Anticlockwise rotate
C: Clockwise rotate
L: Enlarge
S: Small

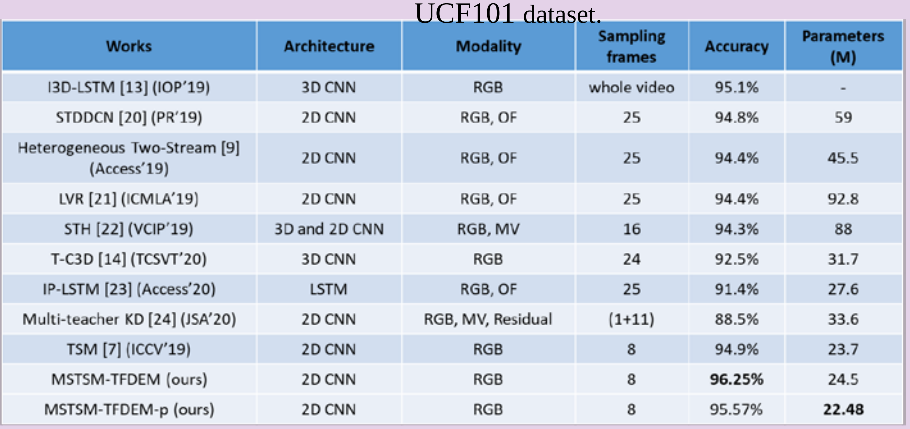


國立清華大學
NATIONAL TSING HUA UNIVERSITY

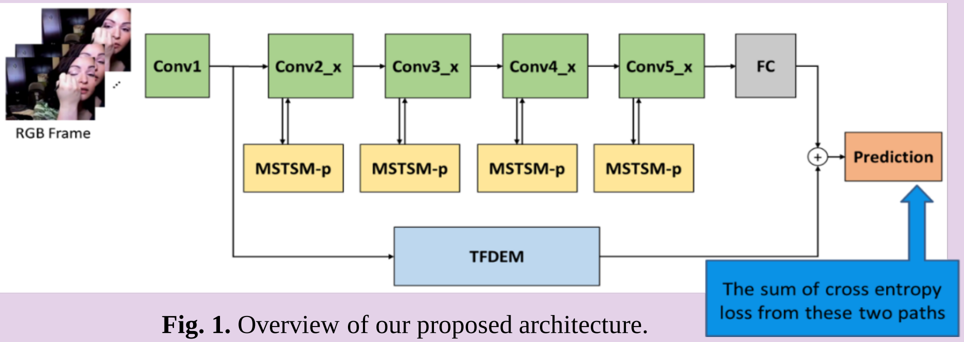
Multi-scale Temporal Shift based 2D CNN for Action Recognition

- ➢ Targets
  - ➢ Propose a framework based on 2D CNN which enlarge the temporal receptive fields with a moderate scale.
  - ➢ Maintain the efficiency.
- ➢ Proposed Solutions
  - ➢ Multi-scale temporal shift module
  - ➢ Temporal feature difference extraction module
  - ➢ Define and prune the similar kernels
- ➢ Contributions
  - ➢ Increase temporal receptive fields by 5x compared with traditional 2D CNN methods.
  - ➢ The two proposed modules improve the accuracy by 1.32% in total on UCF-101 dataset.
  - ➢ The amount of parameters of the proposed model is 22.48M and achieve 95.57% accuracy with inference time of 170fps at TITAN V

Tab. 1. Comparison of accuracy and parameters with the state-of-the-art methods on UCF101 dataset.



Works	Architecture	Modality	Sampling frames	Accuracy	Parameters (M)
I3D-LSTM [13] (IOP'19)	3D CNN	RGB	whole video	95.1%	-
STDDCN [20] (PR'19)	2D CNN	RGB, OF	25	94.8%	59
Heterogeneous Two-Stream [9] (Access'19)	2D CNN	RGB, OF	25	94.4%	45.5
LVR [21] (ICMLA'19)	2D CNN	RGB, OF	25	94.4%	92.8
STH [22] (VCIP'19)	3D and 2D CNN	RGB, MV	16	94.3%	88
T-C3D [14] (TCSVT'20)	3D CNN	RGB	24	92.5%	31.7
IP-LSTM [23] (Access'20)	LSTM	RGB, OF	25	91.4%	27.6
Multi-teacher KD [24] (JSA'20)	2D CNN	RGB, MV, Residual	(1+11)	88.5%	33.6
TSM [7] (ICCV'19)	2D CNN	RGB	8	94.9%	23.7
MSTSM-TFDEM (ours)	2D CNN	RGB	8	**96.25%**	24.5
MSTSM-TFDEM-p (ours)	2D CNN	RGB	8	95.57%	**22.48**



The sum of cross entropy loss from these two paths

Fig. 1. Overview of our proposed architecture.

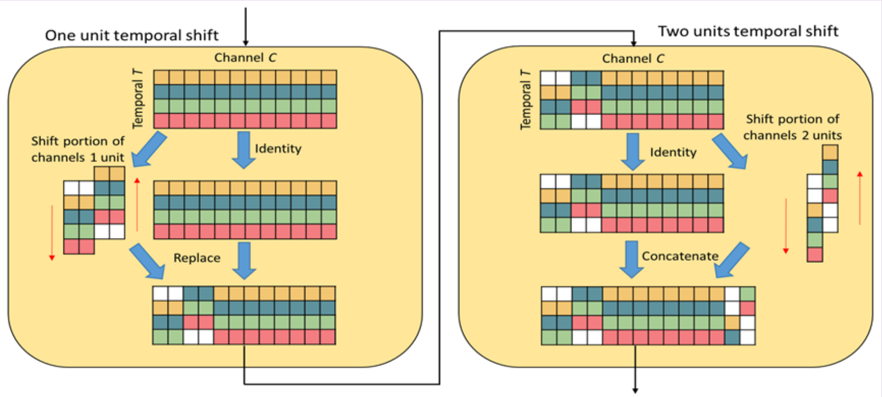


Fig. 2. Architecture of Multi-scale temporal shift.

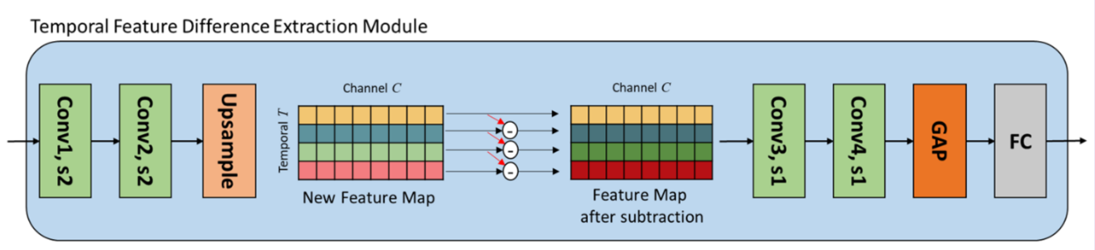


Fig. 3. Architecture of Temporal modeling module.