



工業技術研究院

Industrial Technology
Research Institute

AI System and Application Technology

Tzi-cker Chiueh 闕志克

Information and Communications Labs





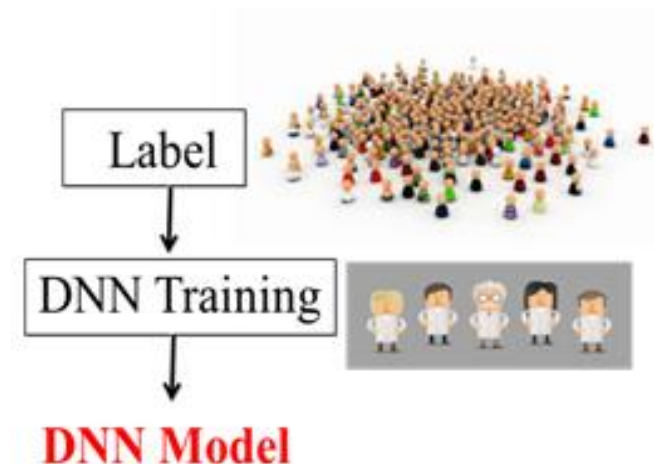
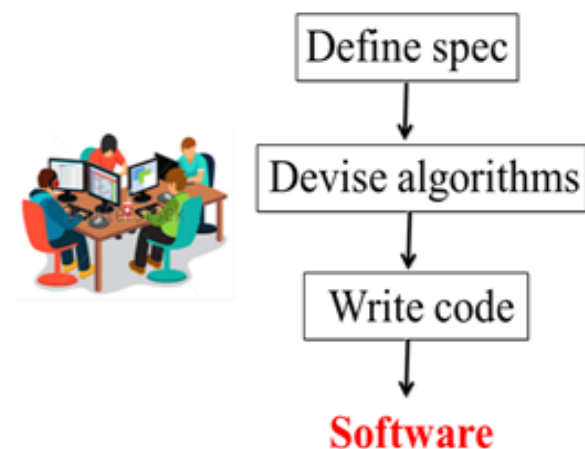
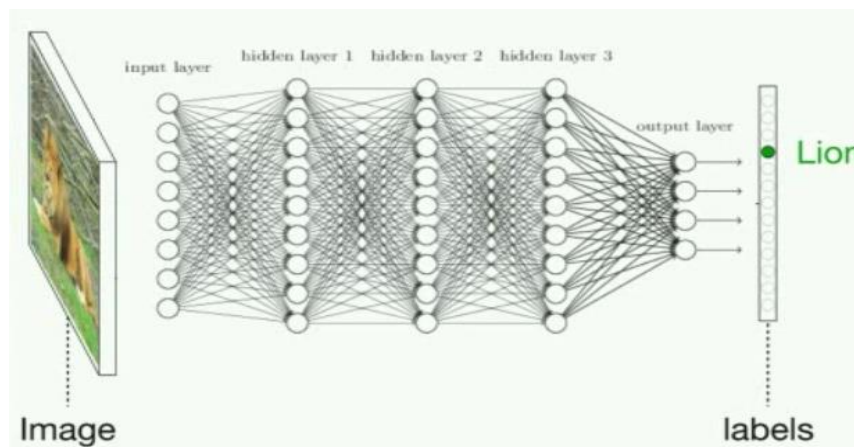
Cost-Effective Private DNN Training System

Why Are DNN Training System Important?

- AI ~ **Deep Neural Network** (DNN)-based Machine Learning
- High-performance DNN training system
 - >30% of the workloads in future data centers
 - **Public DNN training service**: Amazon, Microsoft and Google
 - **Private DNN training appliance**: Nvidia's DGX



Nvidia's DGX-1



Build-up of a DNN Training System

Integrated Development Environment

DNN Training Framework

DNN Model Compiler

Operating System



Training Data → DNN Model

Training Computation to evolve a DNN Model (**TensorFlow, Caffe, and PyTorch**)

DNN Model → Executable Code (**TVM**)

Linux container (**LXC**) and orchestration (**Docker**)

Nvidia/AMD/Intel GPU, FPGA

X86 server with multiple PCIe slots and effective cooling

Open AI Training System Program

- **Goal: Develop an industry for cost-effective DNN training systems**

- **OATS Architecture**

- **Processor Type:**

- Nvidia Tesla P100 and V100 (12GB, 4.7TFLOPs of FP64)
- Nvidia GeForce GTX 1080Ti (11GB, 11.3TFLOPs of FP32)
- AMD Radeon Instinct MI25 (12.3 TFLOPS of FP32)
- AMD Radeon RX Vega 64 (12.6 TFLOPS of FP32)
- Intel multi-node Xeon
- FPGA

- **Graphics driver API:**

- CUDA
- OpenCL

- **DNN training framework:**

- Tensorflow
- Caffe/Caffe2/NVCaffe/PyTorch

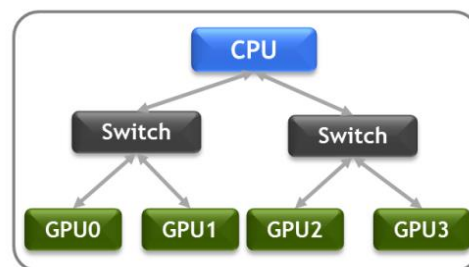
- **Large number of “GPU”s: 16+**

- **System Interconnect:**

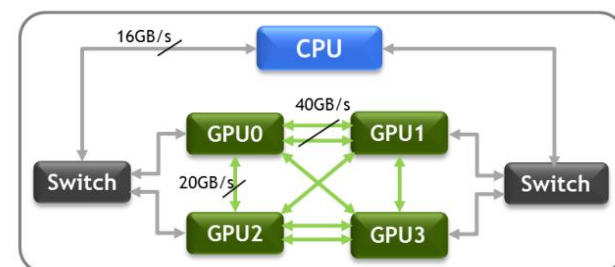
- **Meshed PCIe network**

supporting disaggregate rack architecture

- **Intelligent thermal load management**



4 GPUs with PCIe



4 GPUs with NVLink

Taiwan AI Systems Alliance


- The **TASA** alliance: A technology alliance focused on the development of high-performance DNN training appliances and DNN inference systems targeted at specific AI applications. Members of this alliance include server and industrial computer hardware manufacturers, software stack developers, and users of DNN training and inference systems

Membership List

1. 營邦企業股份有限公司 (Advanced Industrial Computer)
2. 台灣固網股份有限公司 (Taiwan Fixed Networks)
3. 國家實驗研究院國家高速網路與計算中心 (NCHC)
4. 技嘉科技股份有限公司 (Gigabyte)
5. 廣達電腦股份有限公司 (Quanta Computer)
6. 工研院 (ITRI)
7. 凌華科技(股)公司 (Adlink)
8. 宜鼎國際股份有限公司 (Innodisk)
9. 緯穎科技股份有限公司 (Wiwynn)
10. 永擎電子股份有限公司 (ASRock Rack)
11. 英業達股份有限公司 (Inventec)
12. 迎棧科技股份有限公司 (InwinStack)
13. 數位無限軟體股份有限公司 (Infinites Soft)
14. 神雲科技股份有限公司 (MiTAC Computing Technology)



Comparison of DNN Training Systems

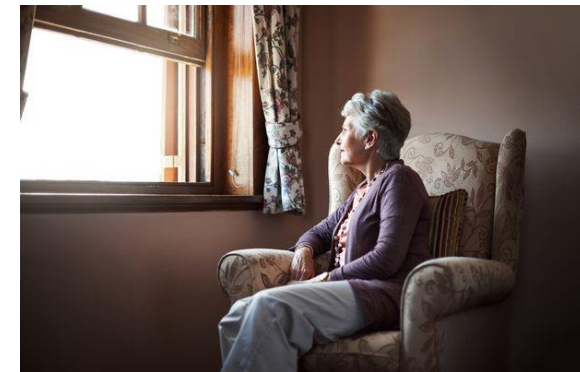
	Vanilla X86+GPU	ITRI DNN Training System 	DGX-1
Hardware	A X86 server + 8 NVIDIA GeForce GTX 1080 Ti GPUs	A X86 server + 8 NVIDIA GeForce GTX 1080 Ti GPUs	8 NVIDIA Tesla V100 SXM2 (NVLink) GPUs
Software	Open Source Suite	ITRI DNN Training Stack	DGX-1 Software (DIGITS)
Price	under USD\$33,000	under USD\$33,000	USD\$149,000
ResNet-50, ILSVRC2012 (1,281,167 images), 30 epochs, NVIDIA Caffe			
Training Time	11h 19m	7h 20m	4h 25m
Accuracy	68.48%	68.86%	67.89%
Images Per Second	946	1,466	2,417
Image Per Second, Per \$USD	0.65X	1X	0.37X
ResNet-50, ILSVRC2012 (1,281,167 images), 30 epochs, TensorFlow			
Training Time	8h 9m	7h 33m	10h 2m
Accuracy	68.32%	70.01%	50.65%
Images Per Second	1,307	1,414	1,063
Image Per Second, Per \$USD	0.92X	1X	0.17X



Companion Robot for Geriatrics

Role of Companion Robot for Geriatrics

- Emerging societal issue: Increasing number of older people living alone (OPLA), because of (1) longer life expectancy and (2) urbanization
- Healthy aging is great for the welfare of individuals and essential for the sustenance of the families and society.
 - In 2017 total care cost of 5.5M dementia patients in US is \$259B
 - Active social engagement and early diagnosis/treatment are two known approaches to combating dementia.
 - Loneliness is a disease and is directly correlated with mortality rate
- Idea: Design a companion robot that helps old people age healthily by getting friends and families involved
- Pain points of children of OPLA
 - Want to but cannot keep close tabs on how OPLAs are doing
 - Have difficulties in engaging meaningfully with loved OPLAs
 - Lack of peace of mind



System Requirements

Design Principles: **No wearable sensor** and **absolute respect for privacy**

- To OPLA

- Electronic photo album



- Personal video communicator



- Multimedia player for entertainment



- Social gaming device



- To friends and families

- Safety emergency alert

- Fall, heart attack, etc.
- Fire, gas leak, break-in, etc.

- Report on living quality

- Sleep well?
- Eat well?
- Feel good?
- Enough exercise?

- Tele-operation & -configuration

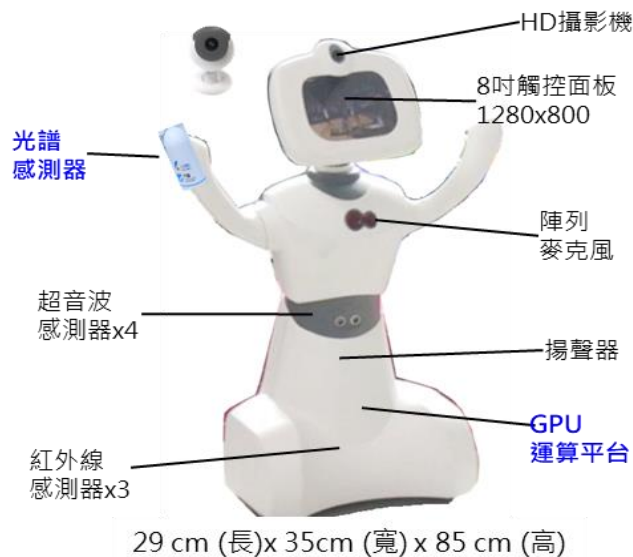
- **Time** and **material** for showing care, concern and attention

AMIBO

Two-Way Video



Tele-Operation and Configuration



*Pecola: Personal Companion for Older People Living Alone

Multimedia Entertainment

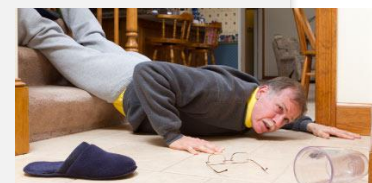
News, movie, TV, and gaming



Social Photo Sharing



Emergency Alert



Exercise Enough?



Eat Well?



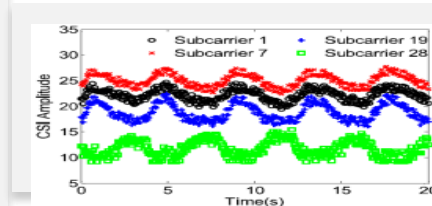
Socializing?



Feel Good?



Sleep Well?

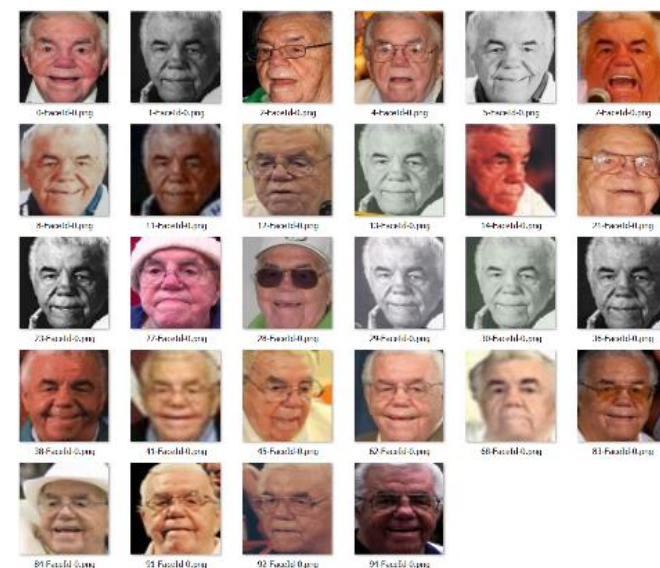


DNN-based Face Recognition

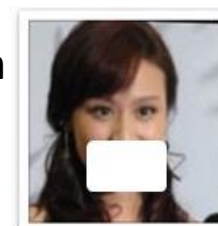
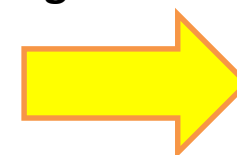
- The world's largest face dataset MS-CELEB-1M, including around 10K persons and 8M images, needs cleaning as different persons assigned the same ID label
- We clean them up and augment the cleaned result



clean



augmentation



Face Detection Performance

Traditional (Boosting) Approach



DNN-based Approach



- ITRI's DNN-based face detection recognition engine could robustly detect faces wearing masks, glasses, and faces at large angles ($>20 \times 20$ pixels)
- Detected faces are passed to the recognition process.

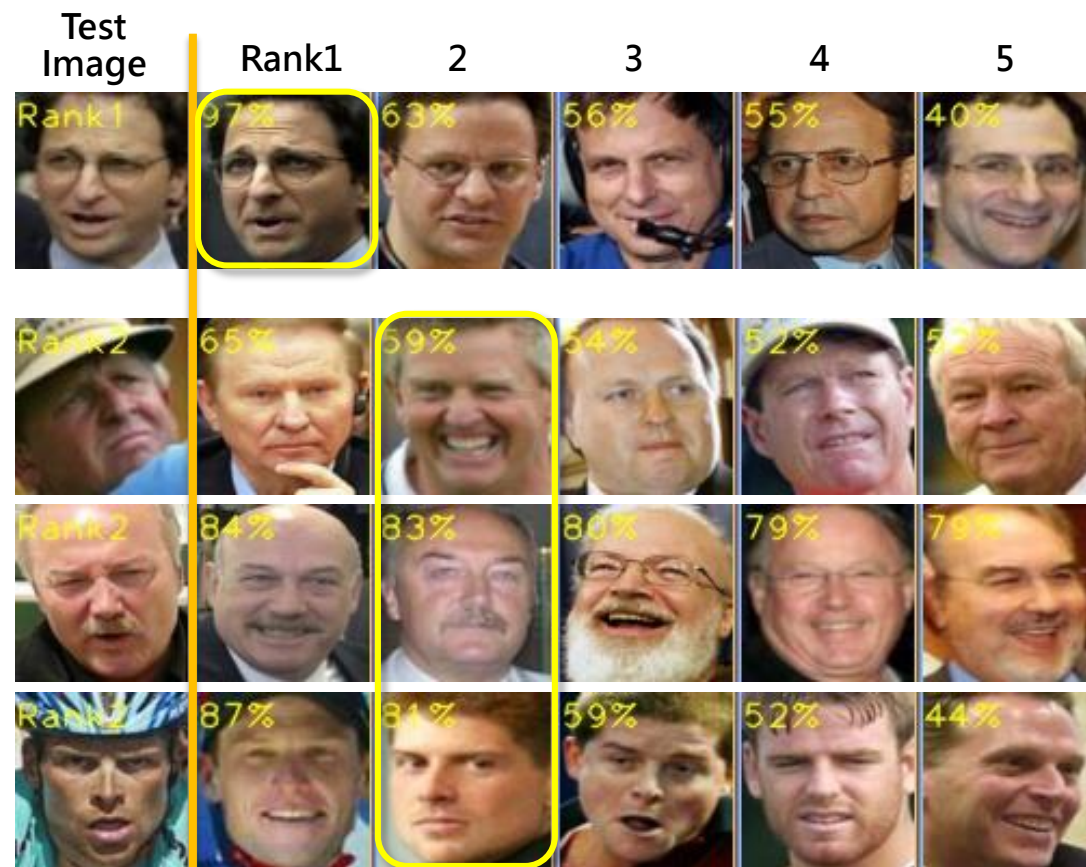
Face Recognition Performance

- LFW+CMU 1,000 persons
- 2 images for training, 1 for test

Rank1	Rank2	Rank3
990	7	1

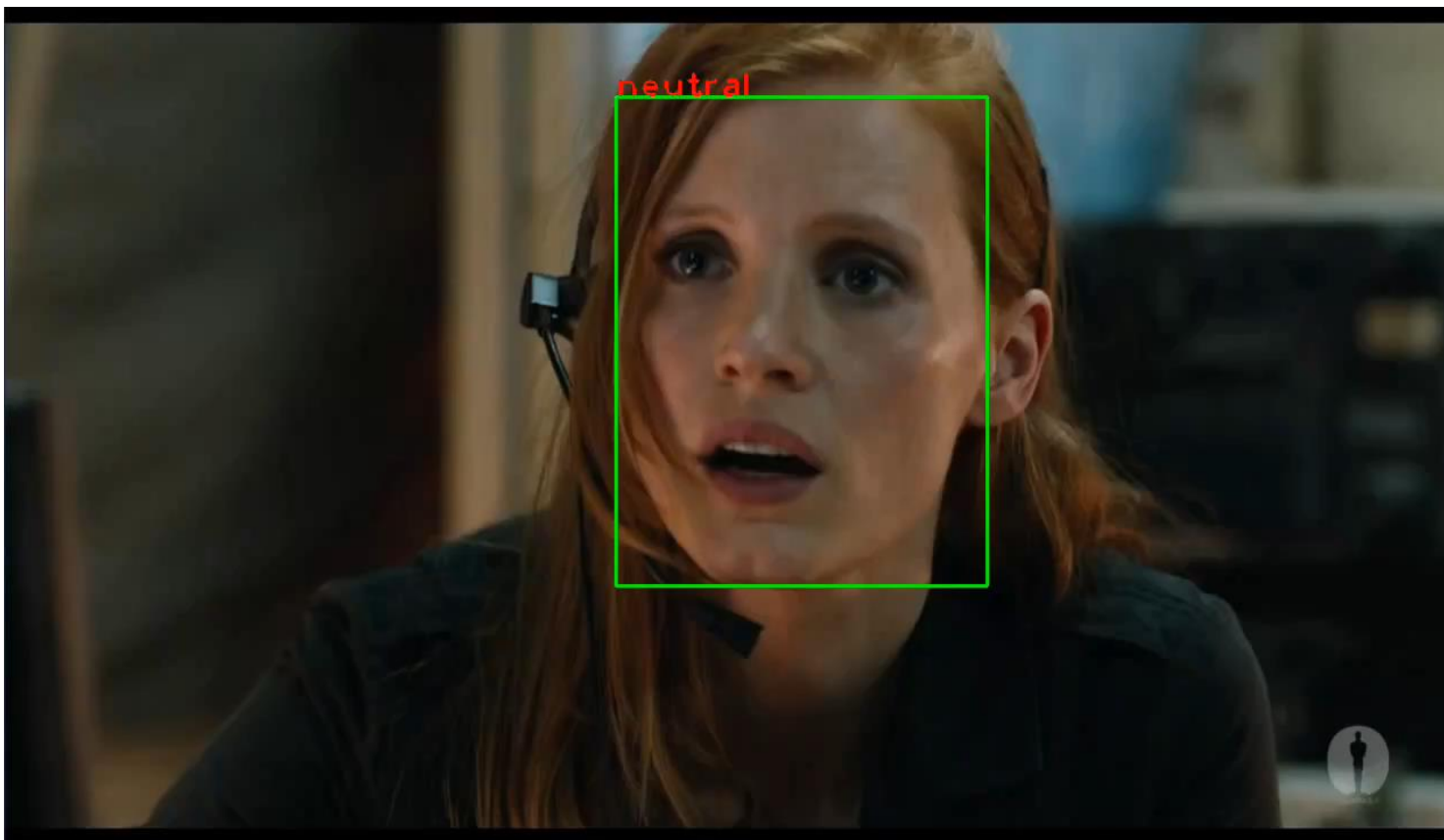
- Time to register a new member (20 images) into the face database: < 5 secs.
- Time to recognize & retrieve Top20 members from a face database:

100,000 Persons	10ms
1,000,000 Persons	60ms
5,000,000 Persons	300ms



Facial Expression Recognition

- ITRI's face expression recognition engine is based on the Resnet architecture and is trained on the FER+ dataset (28385 faces).
- Can recognize 7 emotions, including happiness, sadness, surprise, angry, fear, disgust, and neutral, with an accuracy of 85% and 92% under FER+ and CK+ data sets.
- Speed: Can recognize more than 30 faces and their emotions using a single GTX-1080TI GPU.



Diet Analysis

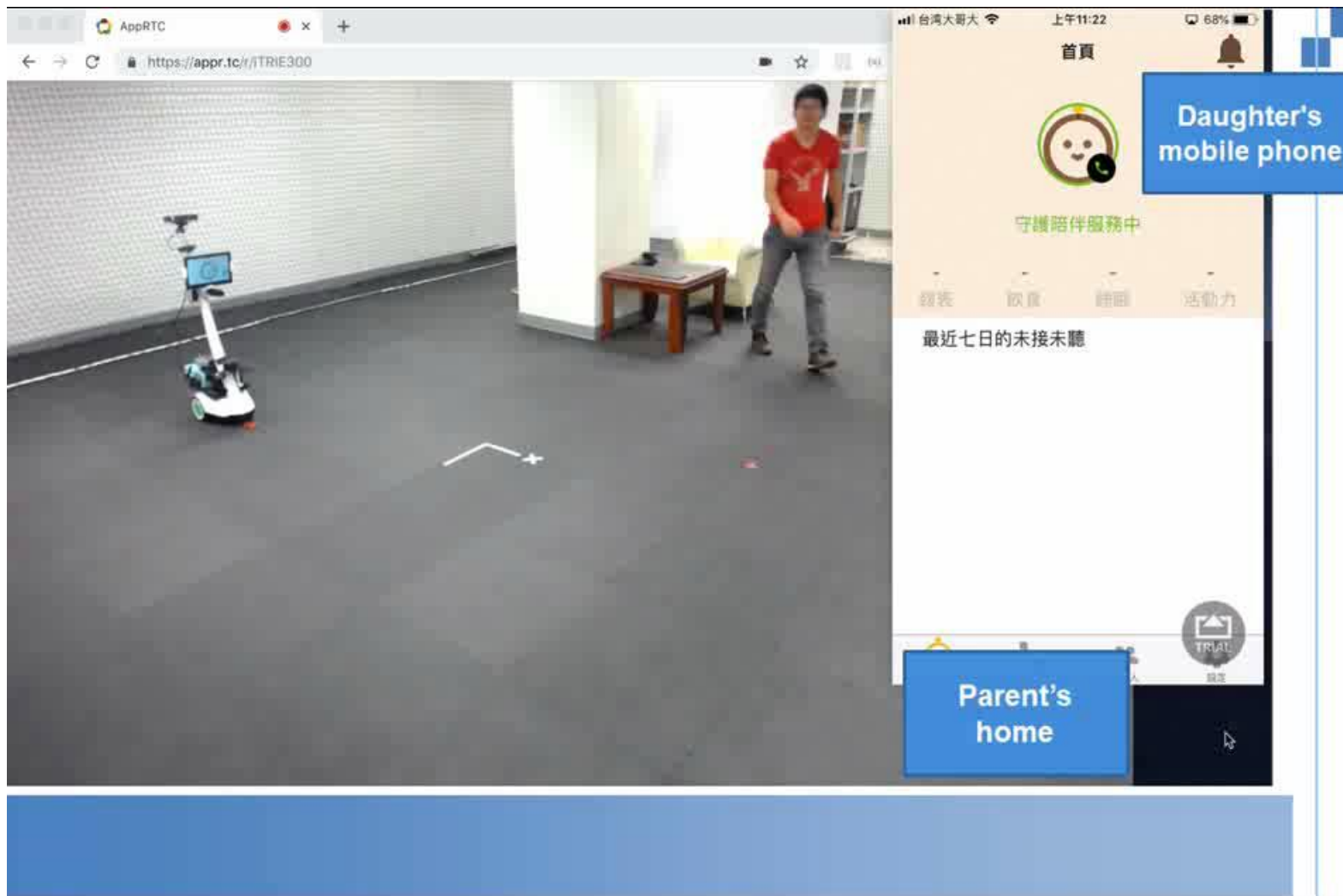
- Food container → food → types of food (rice/pork/fish/vegetable/soup) →
 - Portion change
 - Fresh vs. left-over
 - Specific dish name (Chinese food)
 - Nutrient content



Audio/Video-based Fall Detection

- Some commercial fall detection products require the elderly to wear sensors such as accelerometers, gyroscopes, or level sensors
➔ Inconvenient and false alarm-prone
- ITRI's video –based fall detection engine is based on a training set of 28,776 images each labeled with one of 4 poses (stand/sit/front tilt/lie), and achieves a detection accuracy of 86%.
- In addition, ITRI's fall detection engine includes an audio detection element that could recognize “asking for help” sounds and locate the sound sources ➔ useful for situations that are beyond the reach of video camera, e.g., bathroom

Fall Detection Demo



WiFi as a Sleep Quality Sensor

- Exploit the **CSI (channel state information)** of WiFi/Cellular signals to more accurately measure the body movement of users
 - Indoor positioning
 - Vital signs measurement
- Why choose wireless signals of WiFi/Cellular
 - Nowadays **everyone carries a smartphone all the time**, and all smartphones come with a WiFi/Cellular NIC
 - **WiFi/Cellular CSI-based** body movement measurement
 - ✓ can be independent of the handset operating system and version
 - ✓ does not increase the power consumption of mobile phones
 - ✓ does not affect any mobile phone usage behavior

Centimeter-grade Indoor Positioning

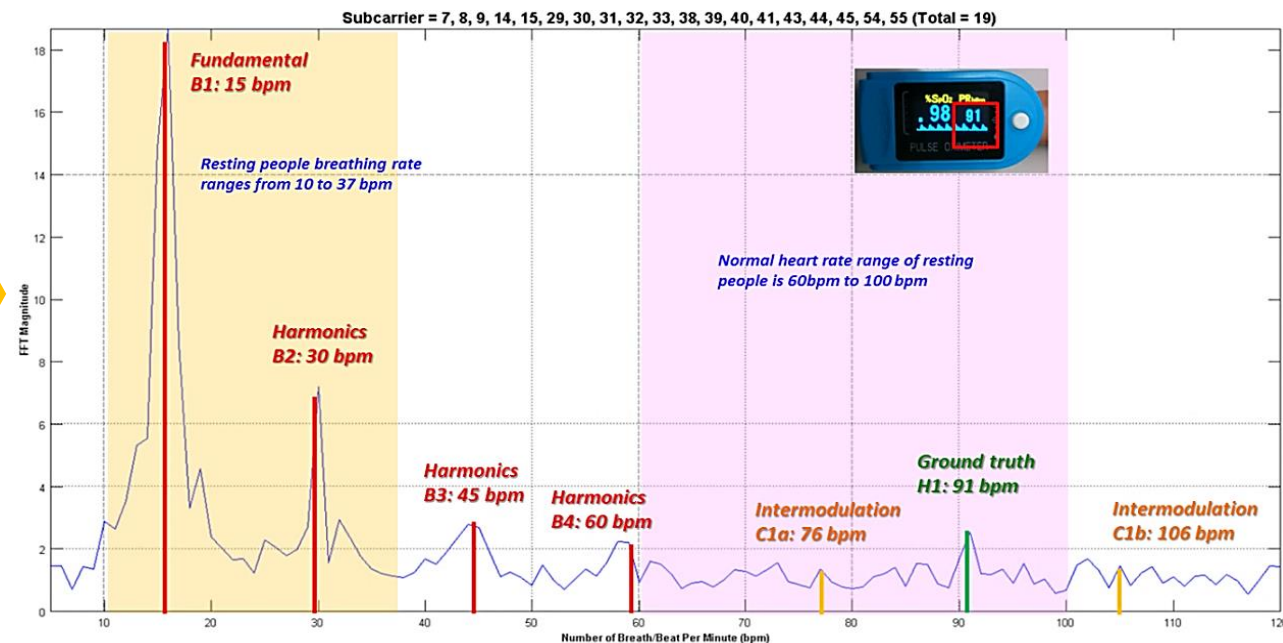
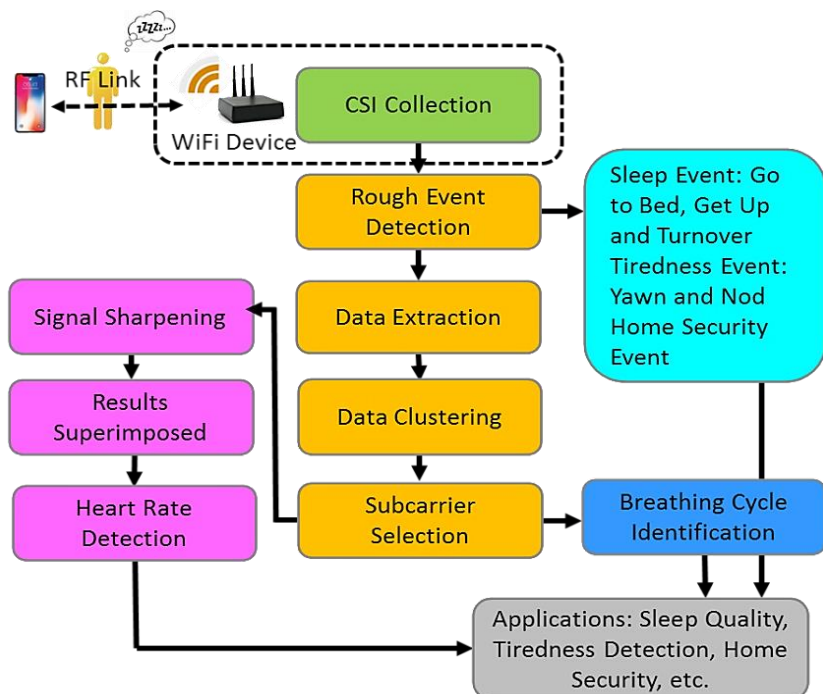
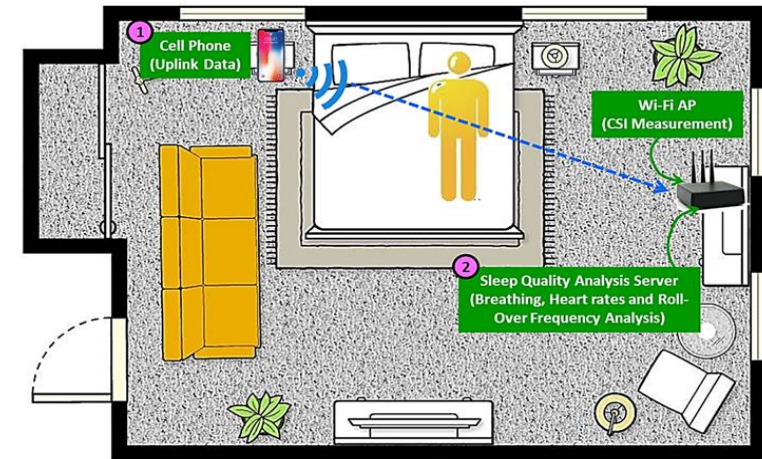


Non-wearable Sleep Quality Detection



WiFi-based Breathing Rate Detection

- Wireless signals are affected by object motion in the environment, including chest movements due to breathing and skin vibrations due to heart beating.
- Apply AI to channel state changes of WiFi signals to inferring minute movements caused by breathing and heart beating
- The accuracy of breathing and heartbeat rate predictions are **90%** and **80%**, respectively, at 1M distance .





WiFi-based Respiratory Rate Detection

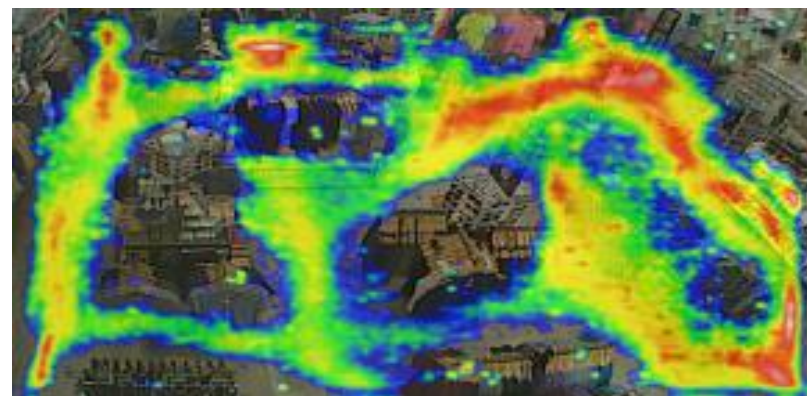
Synchronization signal: 10bpm
Sleep Style: Soldier



Ambient Intelligence for Next- Generation Retail Store

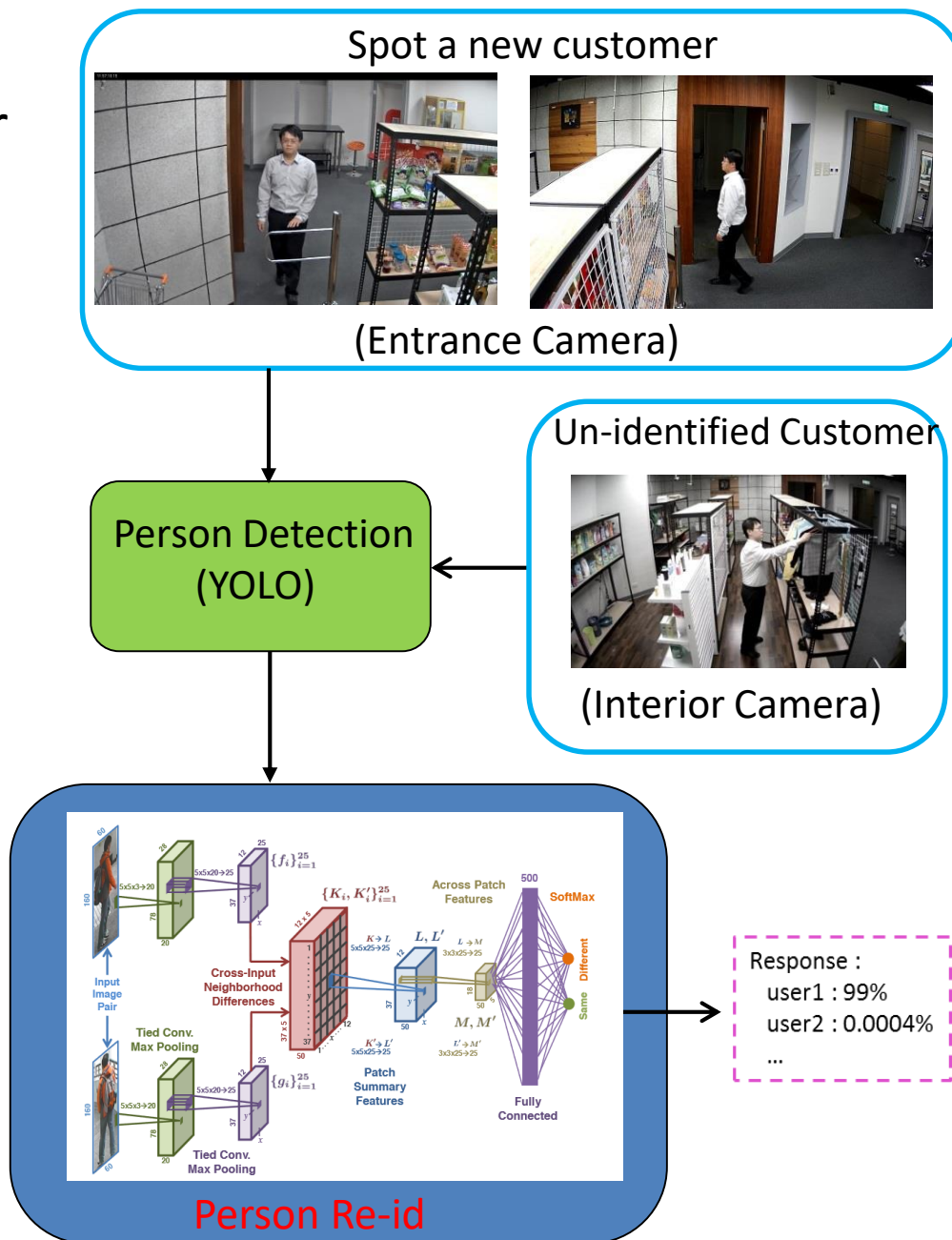
Next-Generation Retail Store

- Design objectives:
 - Optimize shoppers' trial experiences with merchandise
 - Capture the shopper-merchandise interactions as much as possible so as to combine them with on-line shopping behaviors
- Capture whatever can be captured in an E-commerce store in a physical retail store and integrate them across stores
 - How many shoppers are in my store?
 - How does each shopper walk around my store?
 - What merchandise does each shopper interact with?
 - Which merchandise does each shopper like/dislike?



Person Re-Identification

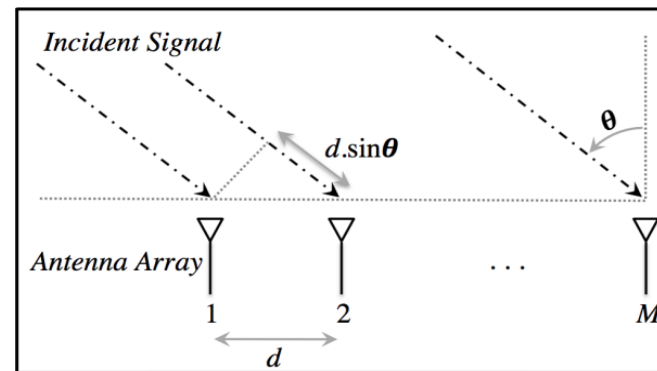
- Use case: Capture each incoming shopper as she walks in, and track her appearances in different cameras
- Binding of person images extracted from multiple cameras
 - A DNN-based method computes a similarity metric for pairs of person images and picks the most similar person for a query person image
 - Person detection → Person image cropping → Person image similarity ranking
 - Accuracy: ITRI's person re-identification engine currently achieves **85%** accuracy under the CUHK03 dataset (1306 persons).
 - Speed: Is able to compare 100 person images per second using a GTX-Titan XP GPU



Centimeter-Resolution WiFi Positioning

Direction of the target

- The channel's **AoA (angle of arrival)** can be used to accurately derive the direction of the target.
- AoA introduces a phase shift across antennas, and phase shift is a function of distance between each antenna and the AoA.

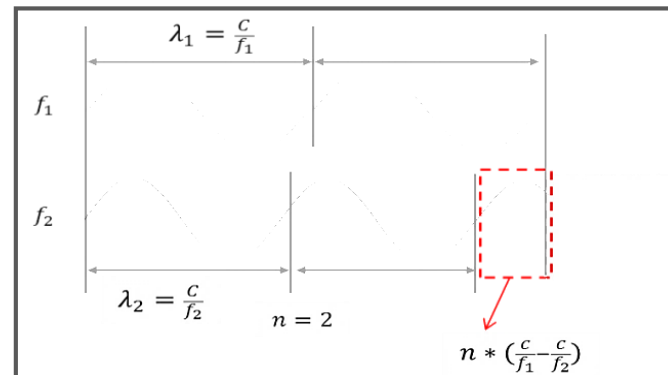


$$\frac{\Delta \text{phase}}{2\pi} = \frac{d \cdot \sin \theta}{\lambda}$$



Distance of the target

- The **received signal strength** can be converted to user distance directly, and signal strength is calculated from the channel's amplitude.
- In addition, we use the **ToF (time of flight)** to tighten up the estimate of the distance.



$$\frac{n * (\frac{c}{f_1} - \frac{c}{f_2})}{\frac{c}{f_2}} = \frac{\Delta}{2\pi}$$

phase shift

$$T_1 = \frac{1}{f_1} \Rightarrow n * T_1 * (f_2 - f_1) \times 2\pi = \Delta$$

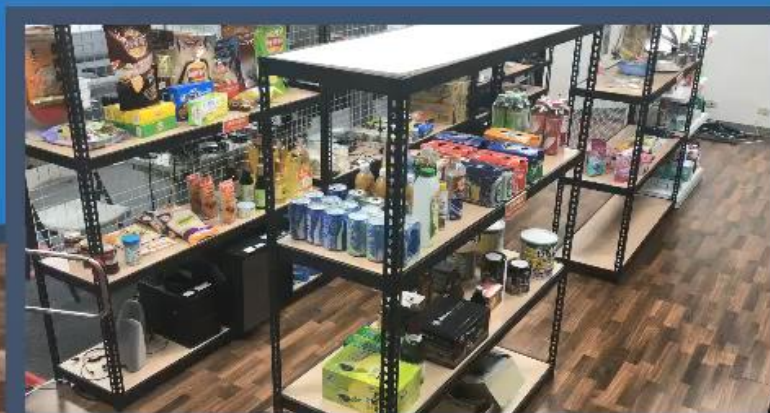
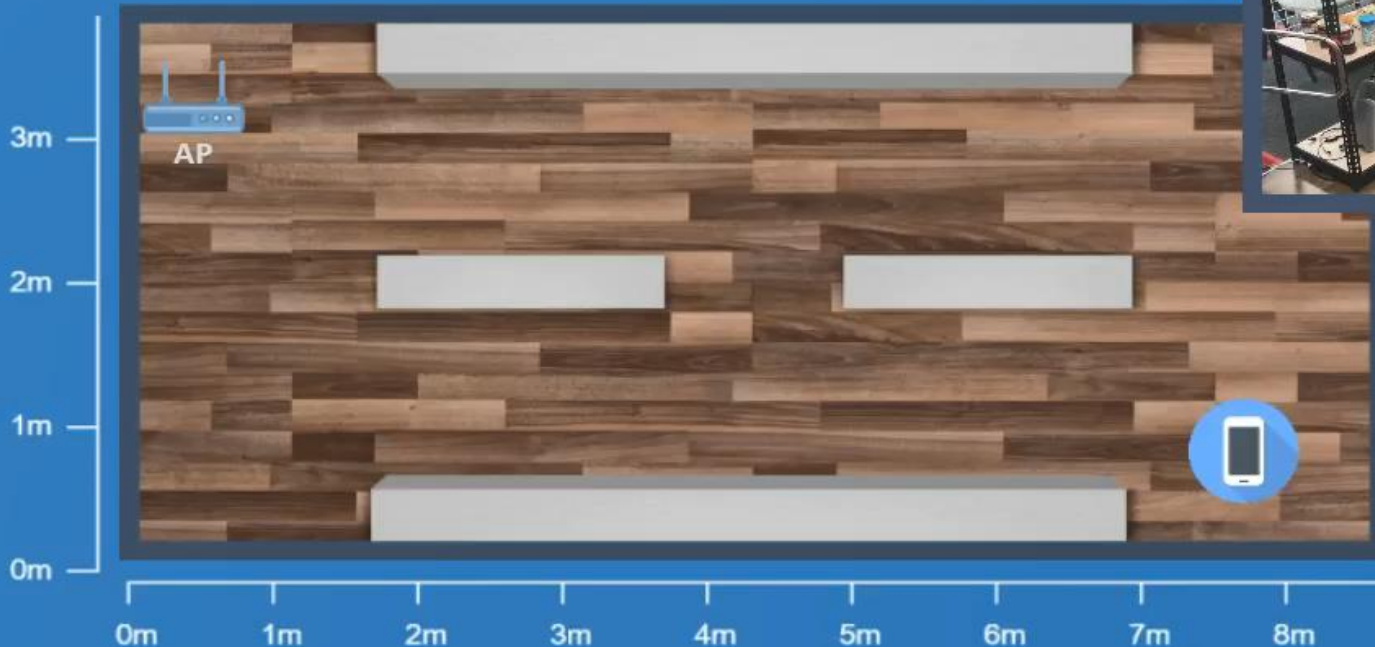
ToF



Centimeter-Resolution WiFi Positioning DEMO

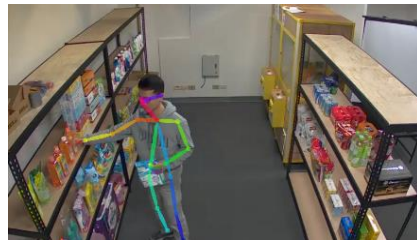


工業技術研究院
Industrial Technology
Research Institute



Shopper-Merchandise Interaction Recognition

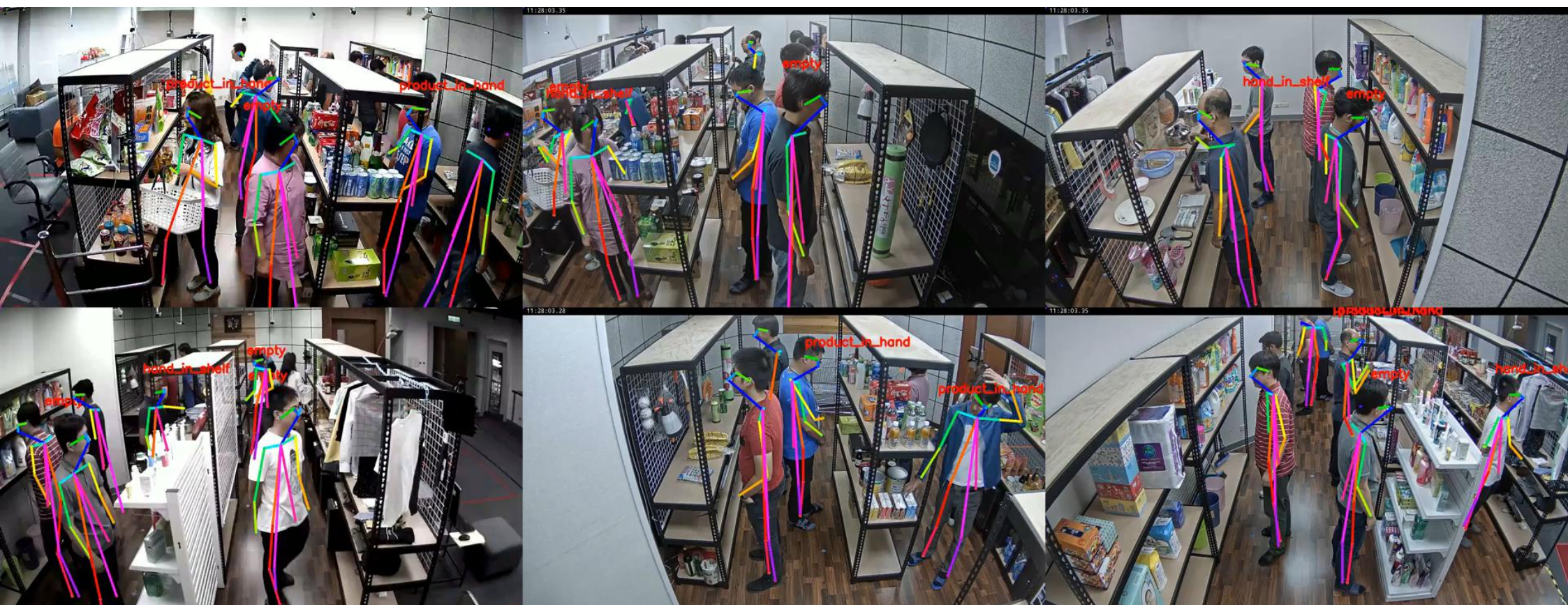
- A two-step approach: (1) a DNN-based **limb movement recognizer** that detects the skeletons of each shopper, and (2) a DNN-based **skeleton-to-action classifier** that deduces how the shopper interacts with the merchandise
- The limb movement recognition engine is trained on the MSCOCO dataset, and achieves a processing speed of 90 frames/second based on GTX-1080Ti.
- The skeleton-to-action classifier, which is based on Resnet and trained on 20,000 labeled images, can recognize seven types of actions, including Empty, Occlusion, Hold Basket, Hold Cart, Hand in Shelf, Product in hand and Put into Cart/Basket, at a recognition accuracy of **77.2%**.
- A pipeline architecture is designed to perform shopper-merchandise interaction recognition in real time.



Video frame	0	1	2	3	4	5				
Video decode										
Limb movement recognition										
Skeleton-to-action inference										
Video encode										
time	0	1	2	3	4	5	6	7	8	9

Shopper-Merchandise Interaction Recognition

DEMO



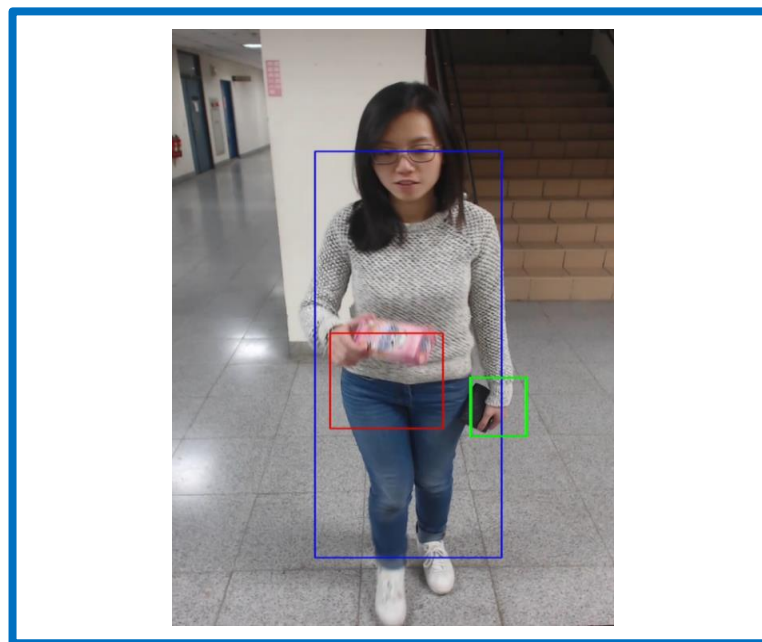
Camera-based Automated Checkout

- Customer places each merchandise in a disjoint area of a plate, with any of its faces facing up
- Particularly useful for convenience stores (e.g. 7-11)

Real-time object recognition



Anomalous behavior detection



Any 3 items of 1000 merchandise, 97% accuracy, ~1 second

Automated Checkout DEMO



Unmanned Open Vending Machine (UOVM)

- **Objective:** An outreach of the 7-11-style convenience store into the commercial work space to cover the last-mile reach
- **Key Idea:** Convert the traditional closed vending machine to **unmanned open vending machine** (UOVM).



Enabling Technologies

- Customer face capturing and recognition (for deterrence)
- Recognition of the merchandise that **customer** picks up
- Inventory capturing after re-stocking by a **staff member**



商品	B01	B02	B03	B04	B05	B06	B07
歌心 歌心歌 糖蜜桃口味	冠寶原味起 司	彩虹糖 彩虹 糖混合水果 口味	健達 健達牛 奶巧克力	雀巢KitKat酥 脆花生奶油 巧克力	雀巢奇巧抹 茶巧克力	Airwaves超 涼黑糖口香 糖-霜芒冰沙 口味	
規格	35(g)	38.2(g)	45(g)	50(g)	38(g)	35(g)	35(g)
售價	\$30	\$30	\$30	\$35	\$39	\$39	\$39
本日銷售量	1	0	0	0	0	1	0
本週銷售量	1	0	0	0	0	1	0
本月銷售量	1	0	0	0	0	1	0
貨架現況影像							



ITRI

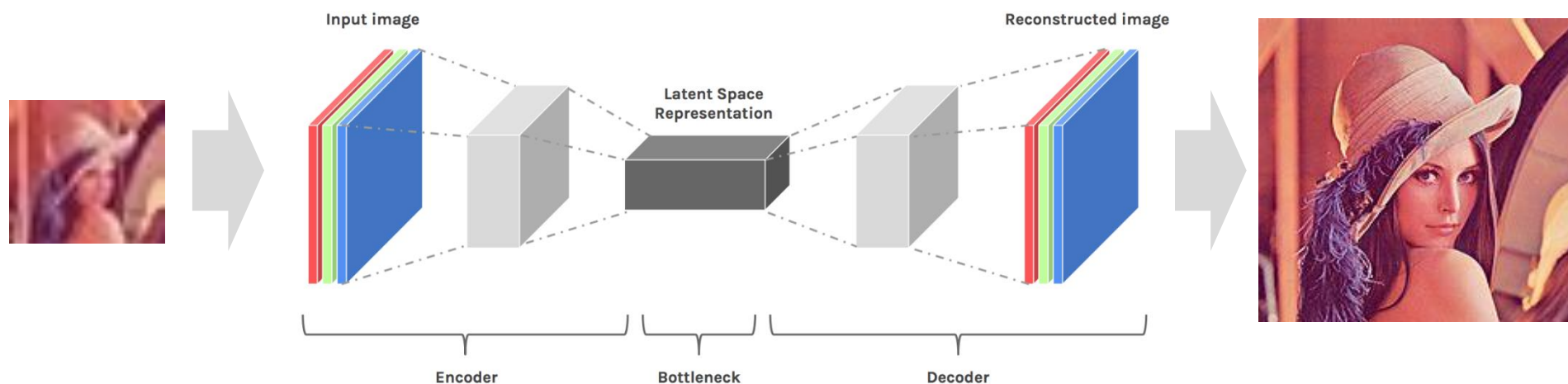
Industrial Technology
Research Institute

Smart Retail Shelf



Super-Resolution (SR)

- Image-to-Image transformation



- Applications:
 - Enlarge the size of an input image with details
 - De-noise and remove artifacts
 - Enrich high frequency details
 - Biomedical image processing

Application to License Plate Recognition

- Percentage of New Taipei City's road-side cameras that become usable increases from **< 20%** to **> 75%**
- Can now recognize license plates of motorcycles



Original Image



0**Q**B8



None

Traditional Bi-cubic
Interpolation



O07B8



0205P**O**

Super-Resolution



0407B8



0205P**B**

Association between Optical and Fluoresce Image

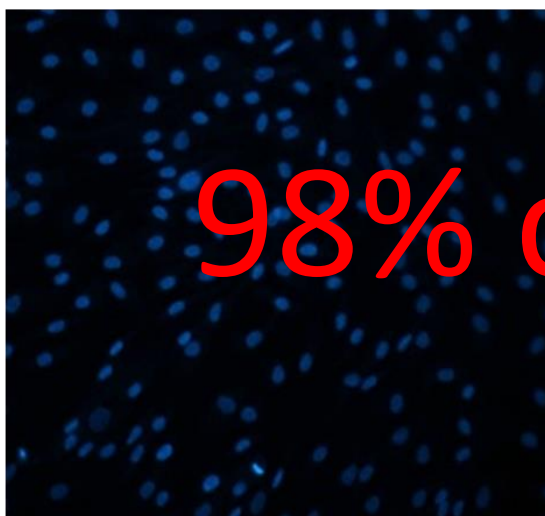
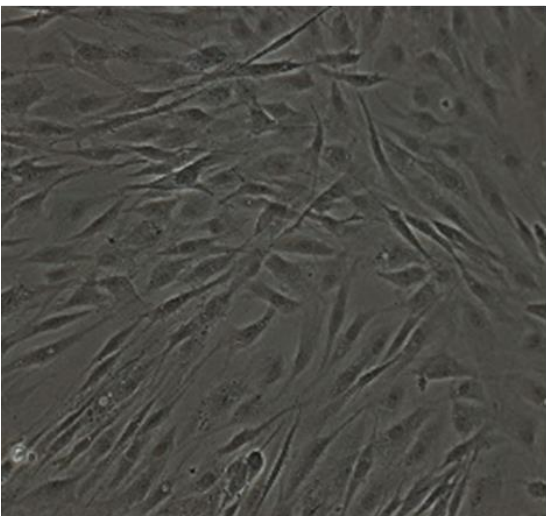
Optical photo

Fluoresce photo



Using the same image-to-image transformation neural network architecture as SR with different datasets

Prediction



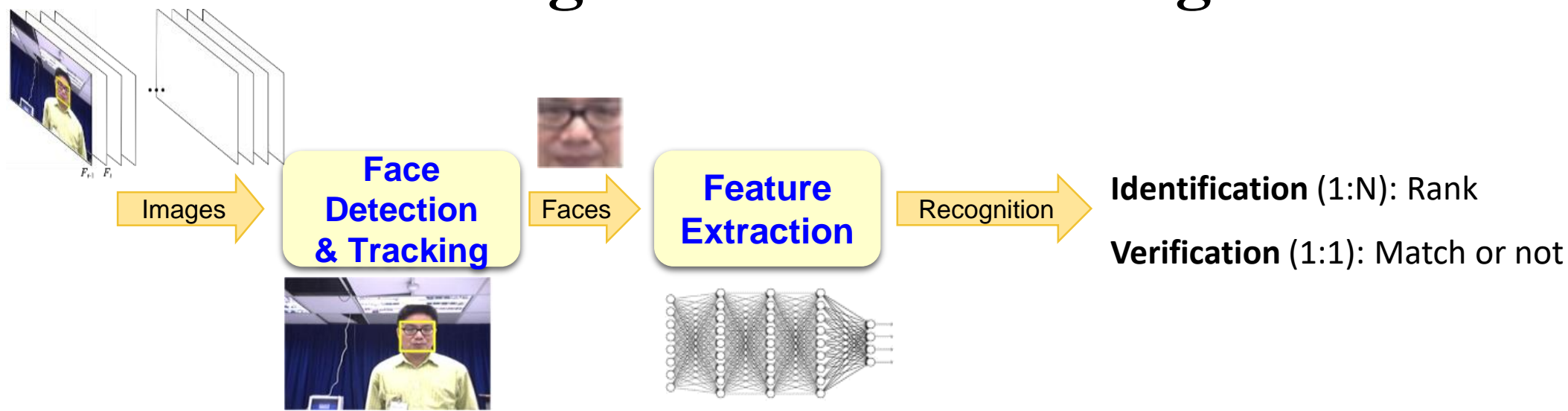


Thank You!

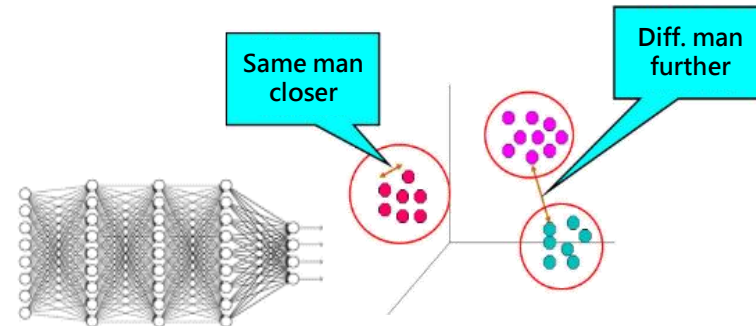
Questions and Comments?

tcc@itri.org.tw

Processing Flow of Face Recognition



Lots of face images (>100M)



extract hundreds features from an image



AMIBO

- To OPLA