



The University of Sydney

# Deep learning for Machine Vision

Wanli Ouyang

wanli.ouyang@sydney.edu.au

# Outline

## Introduction



The University of Sydney

### Our Research Works

Mom and her cute baby are brushing teeth  
Region captioning  
Relationship detection  
Object detection  
Human pose estimation  
Action recognition

### My school

- Computer
- Software
- AI
- Electronics
- Power Engineering
- Renewable Energy
- Smart Micro Grids
- EVs
- Telecommunications
- Wireless remote control
- Sensors
- MEMS
- Fibre Optics

# Outline



## Introduction



The University of Sydney



The University of Sydney

# Deep learning for Machine Vision

Wanli Ouyang

wanli.ouyang@sydney.edu.au



The University of Sydney

# Deep learning for Machine Vision

Wanli Ouyang

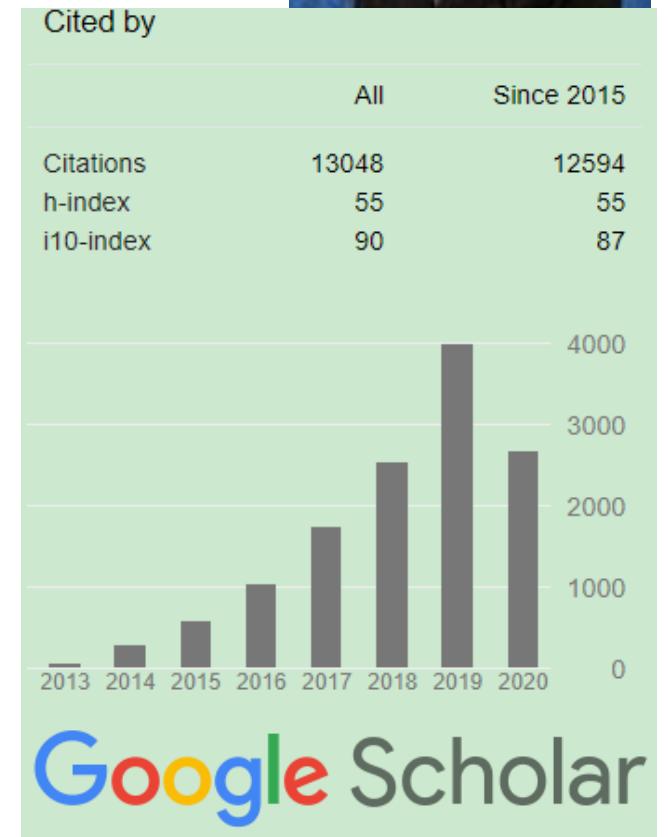
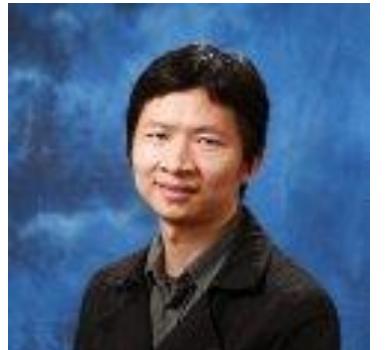
wanli.ouyang@sydney.edu.au

Me

- Deputy head of school on Research
  - School of Electrical and Information Engineering



The University of Sydney



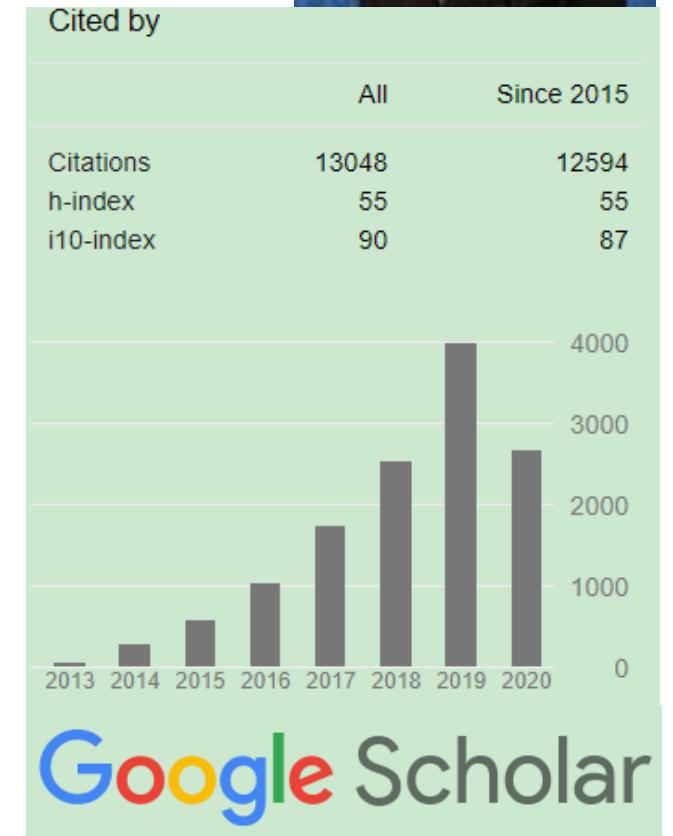
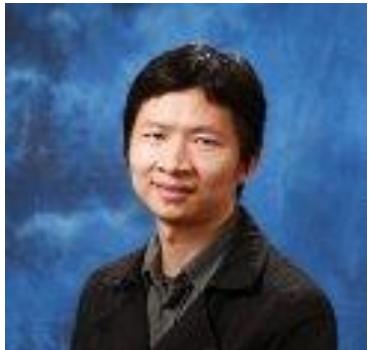
# Me

- Deputy head of school on Research
  - School of Electrical and Information Engineering

Gernot Heiser: a Scientia Professor and the John Lions Chair for operating systems at the UNSW. He is also leader of the Software Systems Research Group at NICTA

Chengqi Zhang: Associate Vice President at the UTS

64	2319		<b>Gernot Heiser</b> UNSW : University of New South Wales Australia	116	12,698	52
65	2336		<b>Wanli Ouyang (欧阳万里)</b> University of Sydney Australia	188	11,827	52
66	2350		<b>Chengqi Zhang</b> University of Technology, Sydney Australia	374	11,070	52





The University of Sydney

# Deep learning for Machine Vision

Wanli Ouyang

wanli.ouyang@sydney.edu.au

# Machine Vision

- Technology and methods used to provide imaging-based automatic inspection and analysis [1].
- Simulate human eyes, and brain
- This talk: mainly about the software part.



[1] [https://en.wikipedia.org/wiki/Machine\\_vision](https://en.wikipedia.org/wiki/Machine_vision)

# Explosive growth in visual data



Hundreds of millions of Videos on YouTube

Youtube

~ 144 million in 2008



# Explosive growth in visual data

Unlimited visual data from IoT, Home Videos, Smart phones, Games ...

Tens of Billions of Photos on Facebook

Hundreds of millions of Videos on YouTube



Facebook

**300 million** – photos uploaded everyday

**15+ billion** – photos on 2008



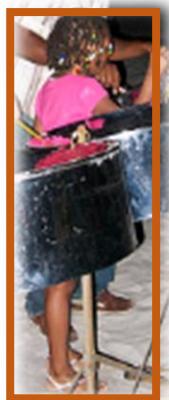
# Challenges -- person

- Variation
  - Color



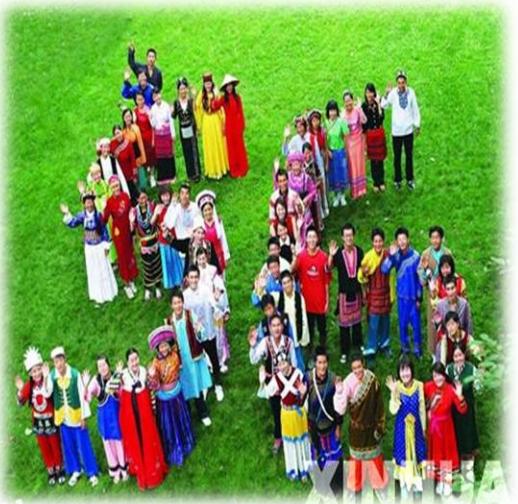
# Challenges -- person

- Variation
  - Color
  - Occlusion



# Challenges -- person

- Variation
  - Color
  - Occlusion
  - Deformation



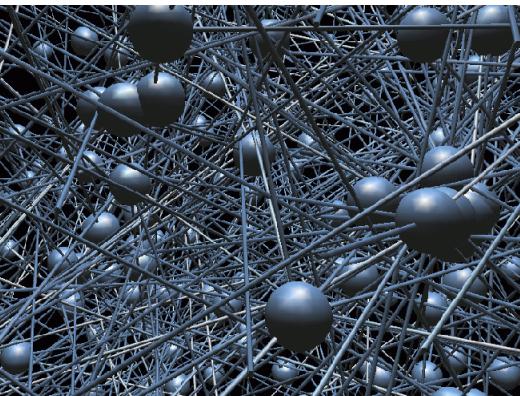


The University of Sydney

# Deep learning for Machine Vision

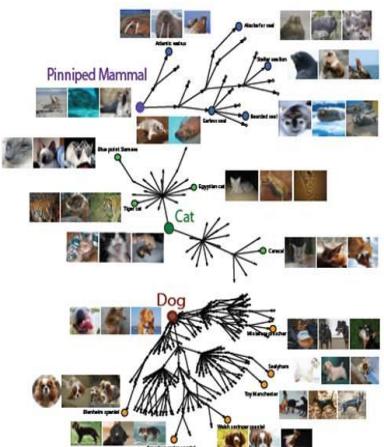
Wanli Ouyang

[wanli.ouyang@sydney.edu.au](mailto:wanli.ouyang@sydney.edu.au)



**Simulate brain activities and employ millions of neurons to fit billions of training samples. Deep neural networks are trained with GPU clusters with tens of thousands of processors**

Hinton won ImageNet competition  
Classify 1.2 million images into 1,000 categories  
**Beating existing computer vision methods by 20%**  
Surpassing human performance



## Deep learning

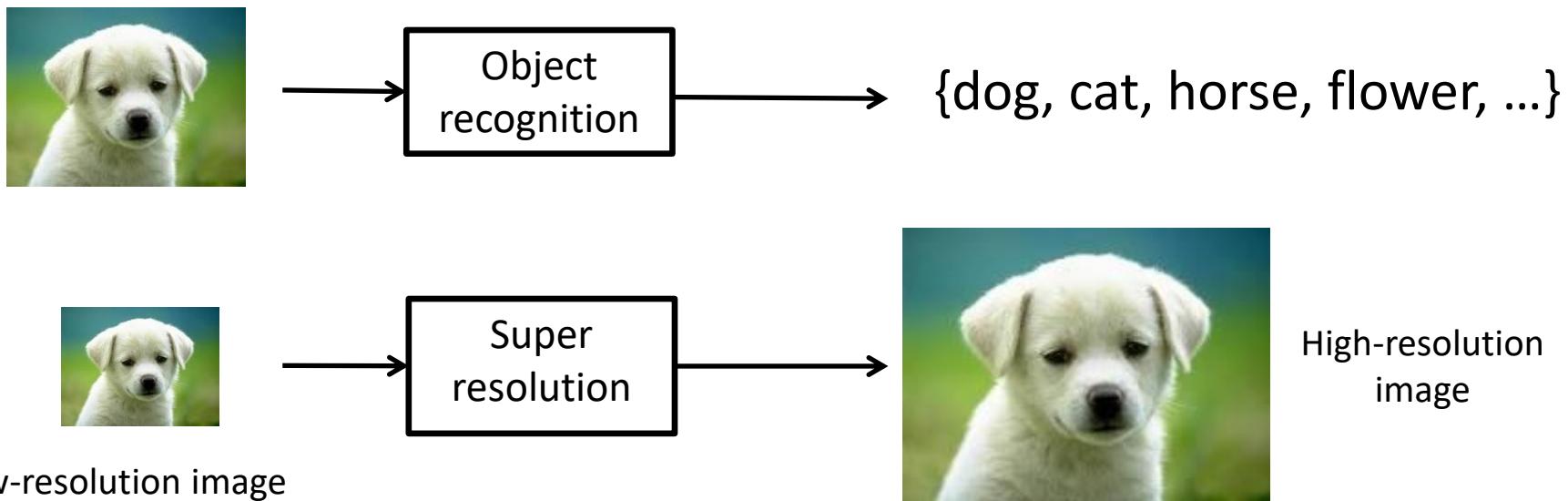
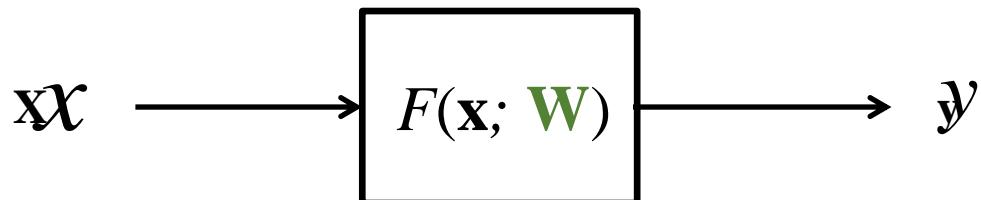
**REVOLUTIONARY**  
Web-scale visual search,  
self-driving cars,  
surveillance, multimedia  
...  
**Hold records on most of the computer vision problems**

**MIT Tech Review**  
**Top 10 Breakthroughs 2013**  
**Ranking No. 1**

**DeepLearning**

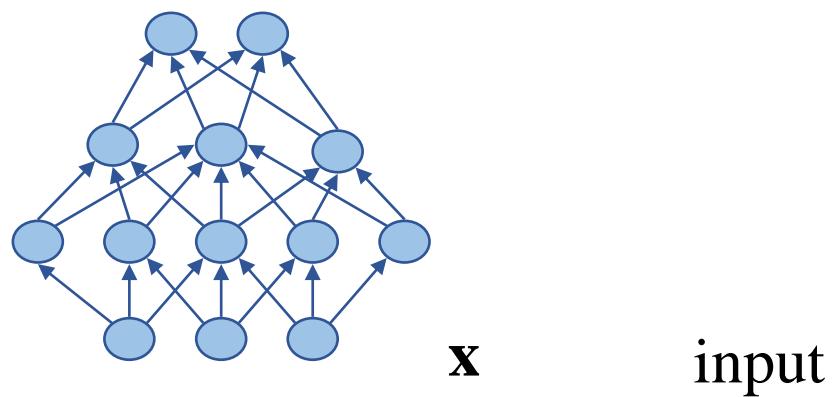
With massive amounts of computational power, machines can now recognize objects and translate speech in real time. Artificial intelligence is finally getting smart.

# Machine Learning



# Deep learning

- Learning of multi-layer non-linear transformations

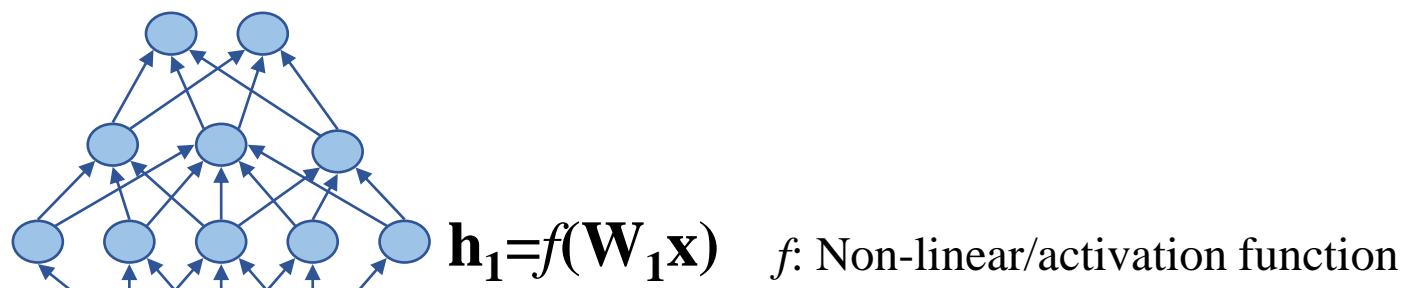


$\mathbf{x}$

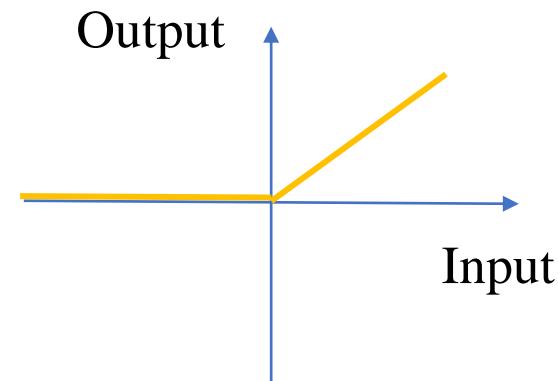
input

# Deep learning

- Learning of multi-layer non-linear transformations

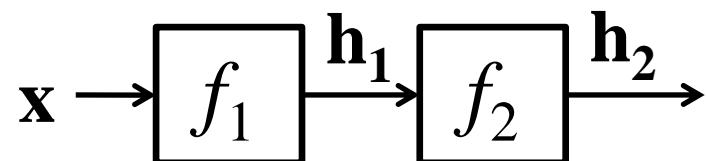
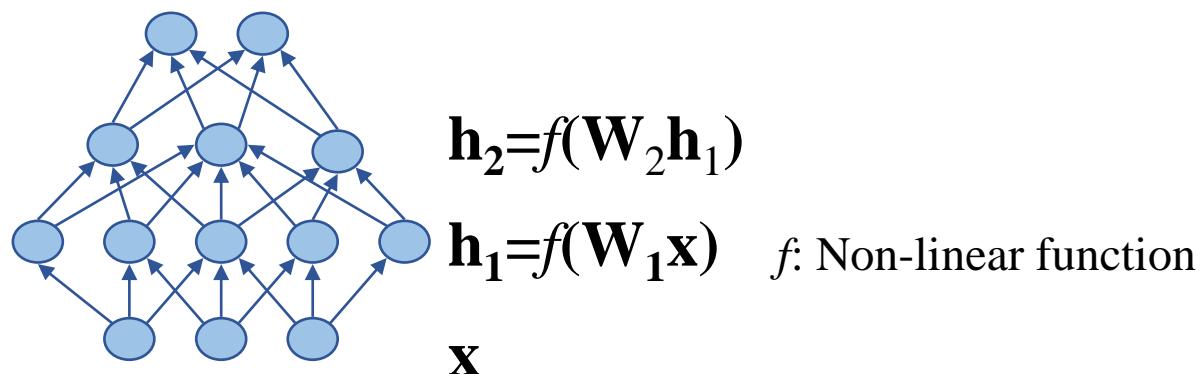


$$\mathbf{x} \rightarrow \boxed{f_1} \quad \mathbf{h}_1$$



# Deep learning

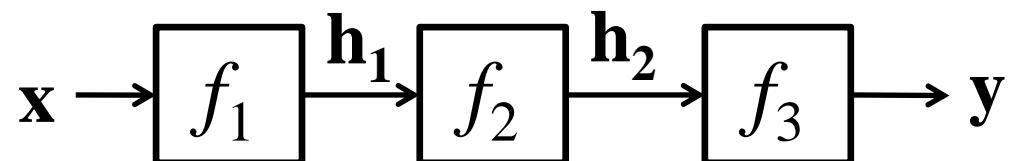
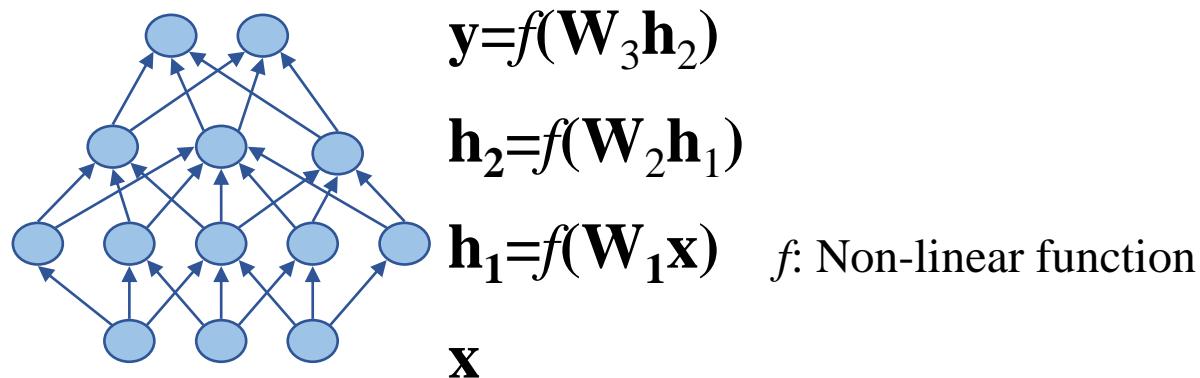
- Learning of multi-layer non-linear transformations



# Deep learning

- Learning of multi-layer non-linear transformations

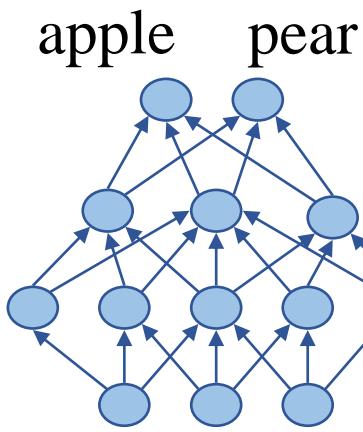
$\mathbf{W} = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3\}$ : parameters



# Deep learning

- Learning of multi-layer non-linear transformations

$\mathbf{W} = \{\mathbf{W}_1, \mathbf{W}_2, \mathbf{W}_3\}$ : parameters

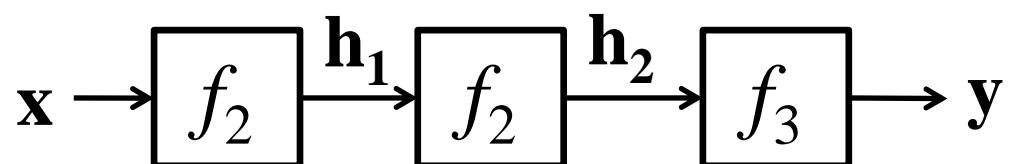


$$y = f(\mathbf{W}_3 h_2)$$

$$h_2 = f(\mathbf{W}_2 h_1)$$

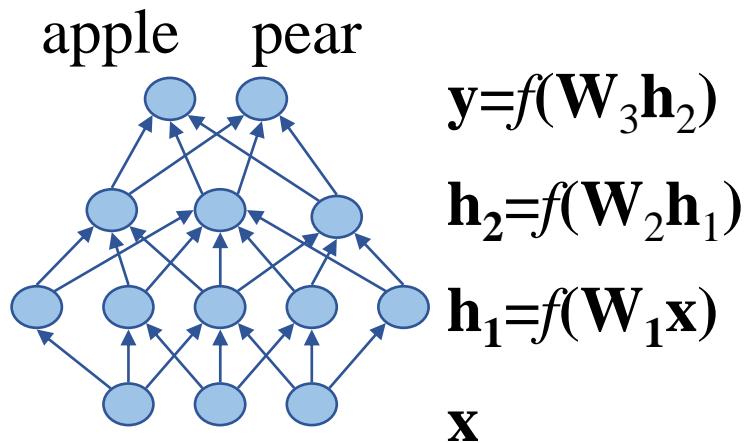
$$h_1 = f(\mathbf{W}_1 x)$$

apple      pear



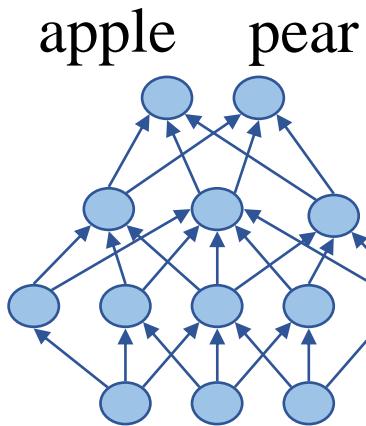
# Deep learning

- Learning of multi-layer non-linear transformations
- Multiple layers have multiple levels of representations with multiple non-linear transformations



# Deep learning

- Learning of multi-layer non-linear transformations
- Multiple layers have multiple levels of representations with multiple non-linear transformations
- Higher level representations are derived from lower level representations to form a hierarchical representation.
- Multiple levels of representations correspond to different levels of abstraction; the levels form a hierarchy of concepts.



$$\begin{aligned}y &= f(\mathbf{W}_3 \mathbf{h}_2) \\ \mathbf{h}_2 &= f(\mathbf{W}_2 \mathbf{h}_1) \\ \mathbf{h}_1 &= f(\mathbf{W}_1 \mathbf{x})\end{aligned}$$

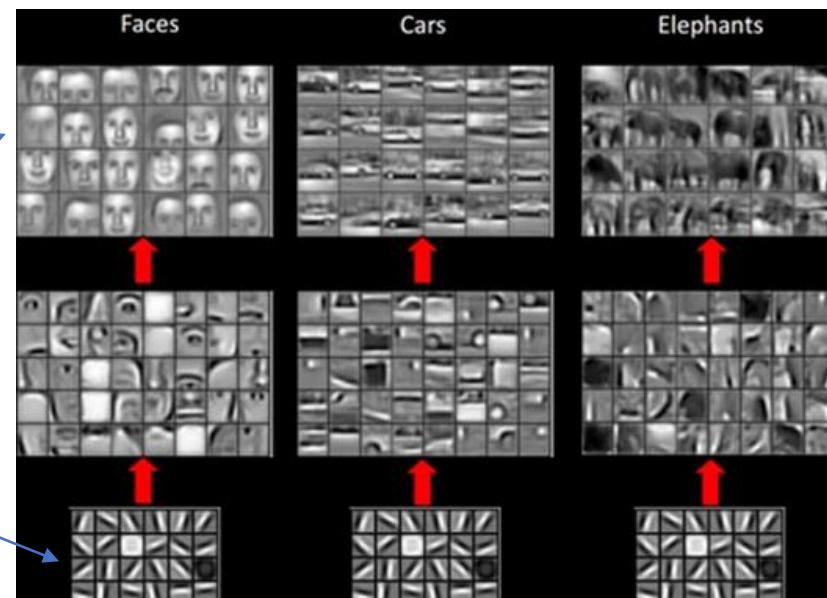


Image from Andrew Ng's Slides

# Deep learning

- Turing award (the ‘Nobel Prize’ in computer and information engineering)
- Hinton, LeCun & Bengio Named 2018 ACM Turing Award Laureates, because of their contribution to deep learning.



# Outline

## Introduction



# Outline



# Outline

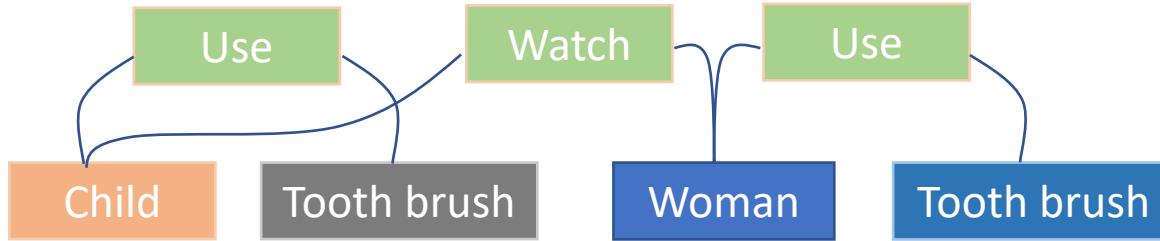
Introduction

Our Research Works

Research topics



Mom and her cute baby are brushing teeth

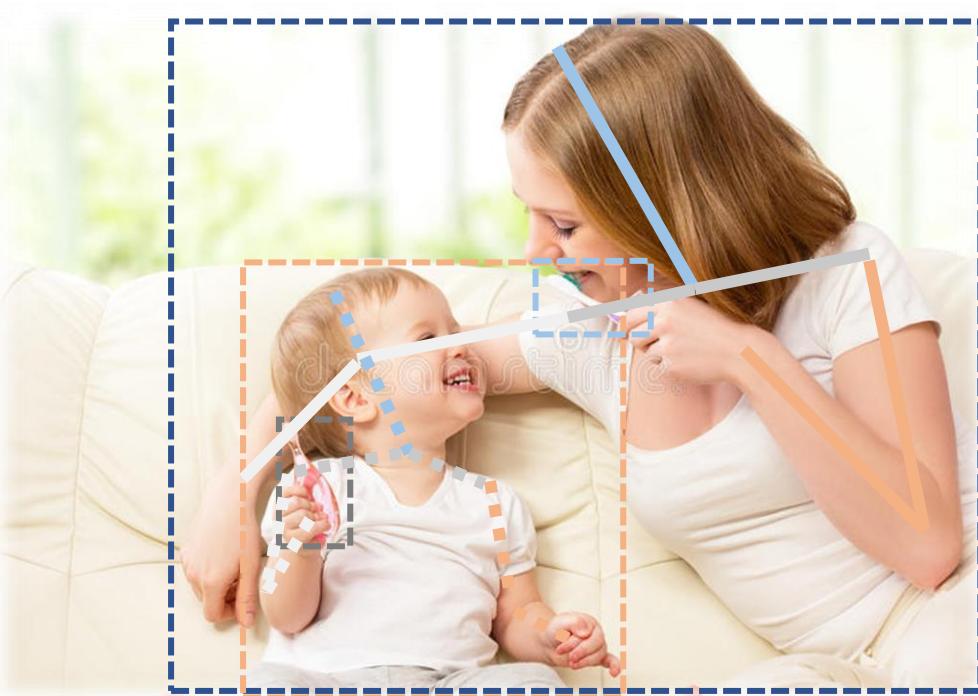


Region captioning

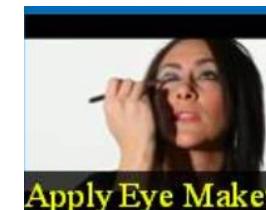
Relationship detection

Object detection

Human pose estimation



Action recognition

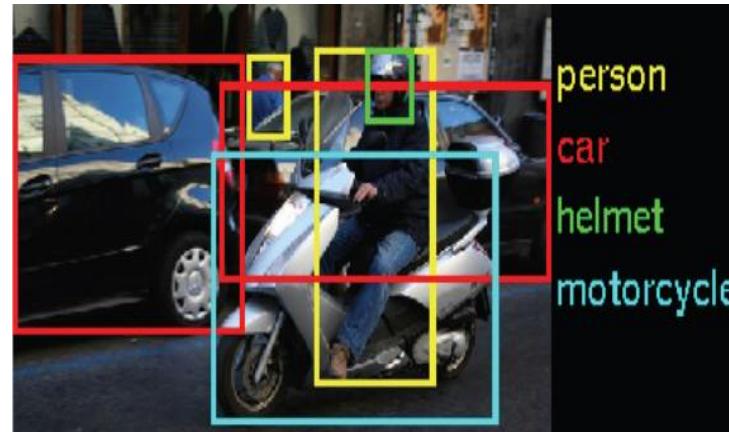
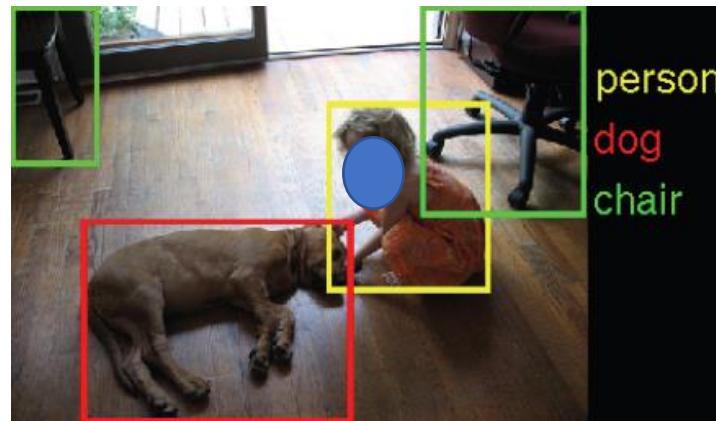
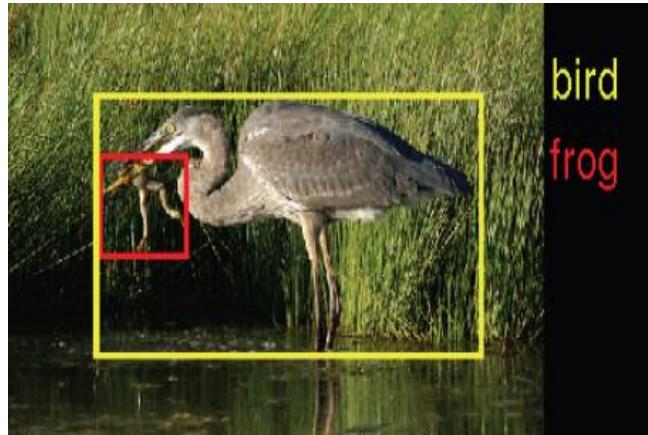


# ImageNet Large Scale Visual Recognition Challenge

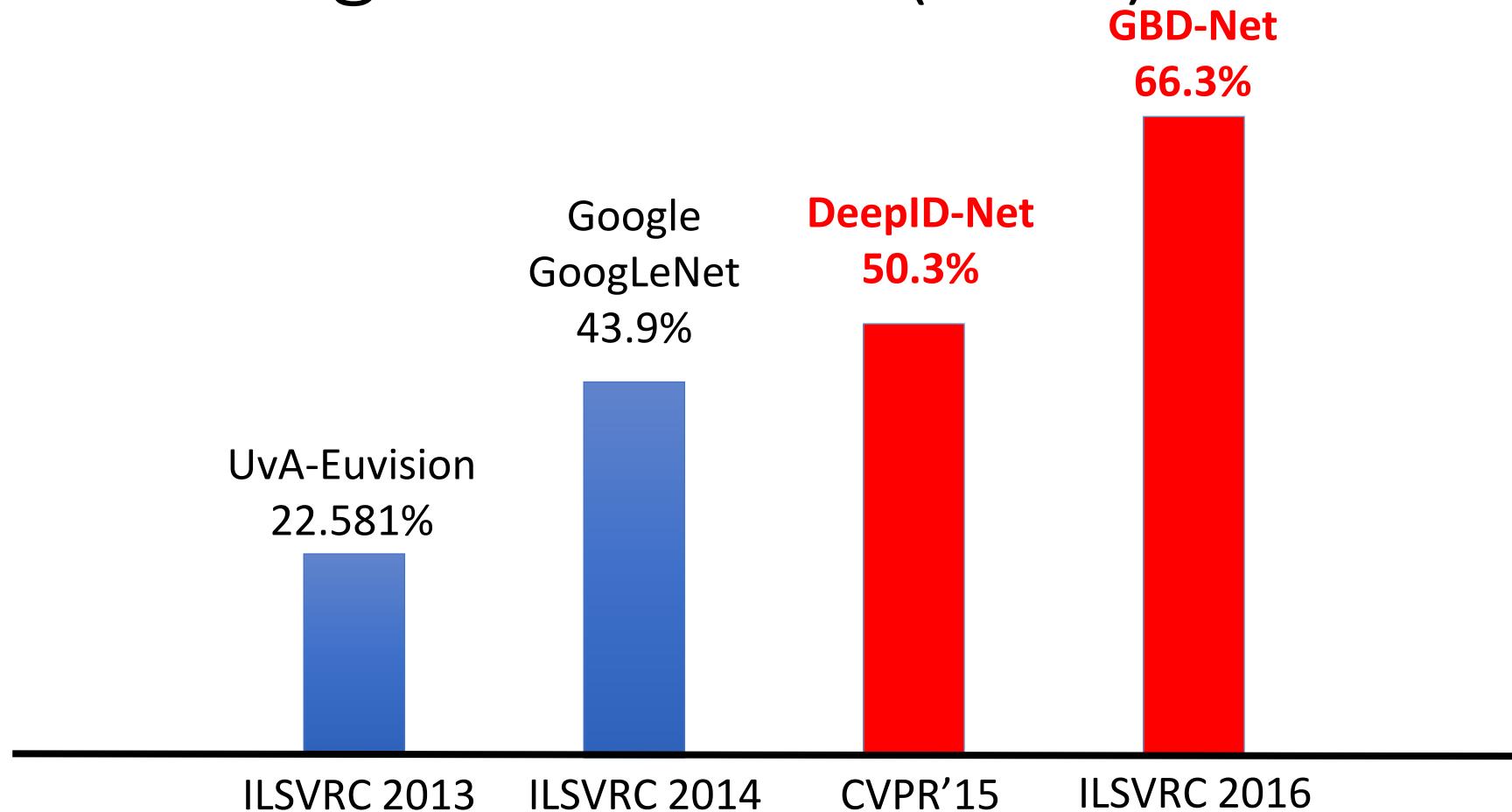


# ImageNet Object Detection Task

- 200 object classes
- ~500,000 training images, 60,000 test images



# Mean Averaged Precision (mAP)



W. Ouyang and X. Wang, et al. "DeepID-Net: Deformable Deep Convolutional Neural Networks for Object Detection," CVPR15, TPAMI17

X. Zeng, W. Ouyang, J. Yan, etc, "Crafting gbd-net for object detection," ECCV16, TPAMI 2017

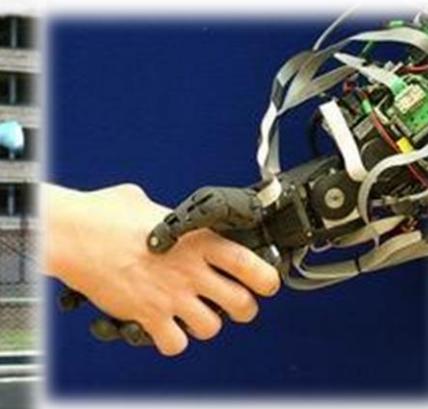
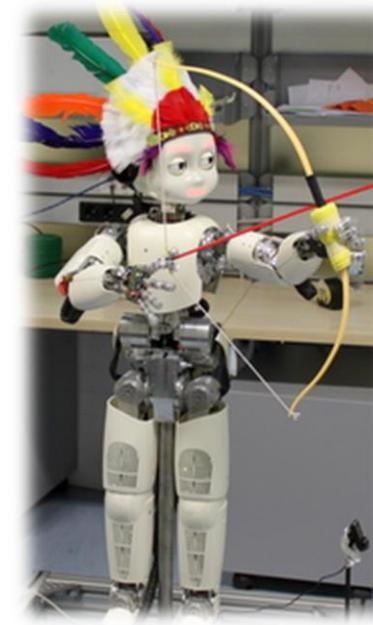
# Application

- ▶ Automotive safety and automatic car driving



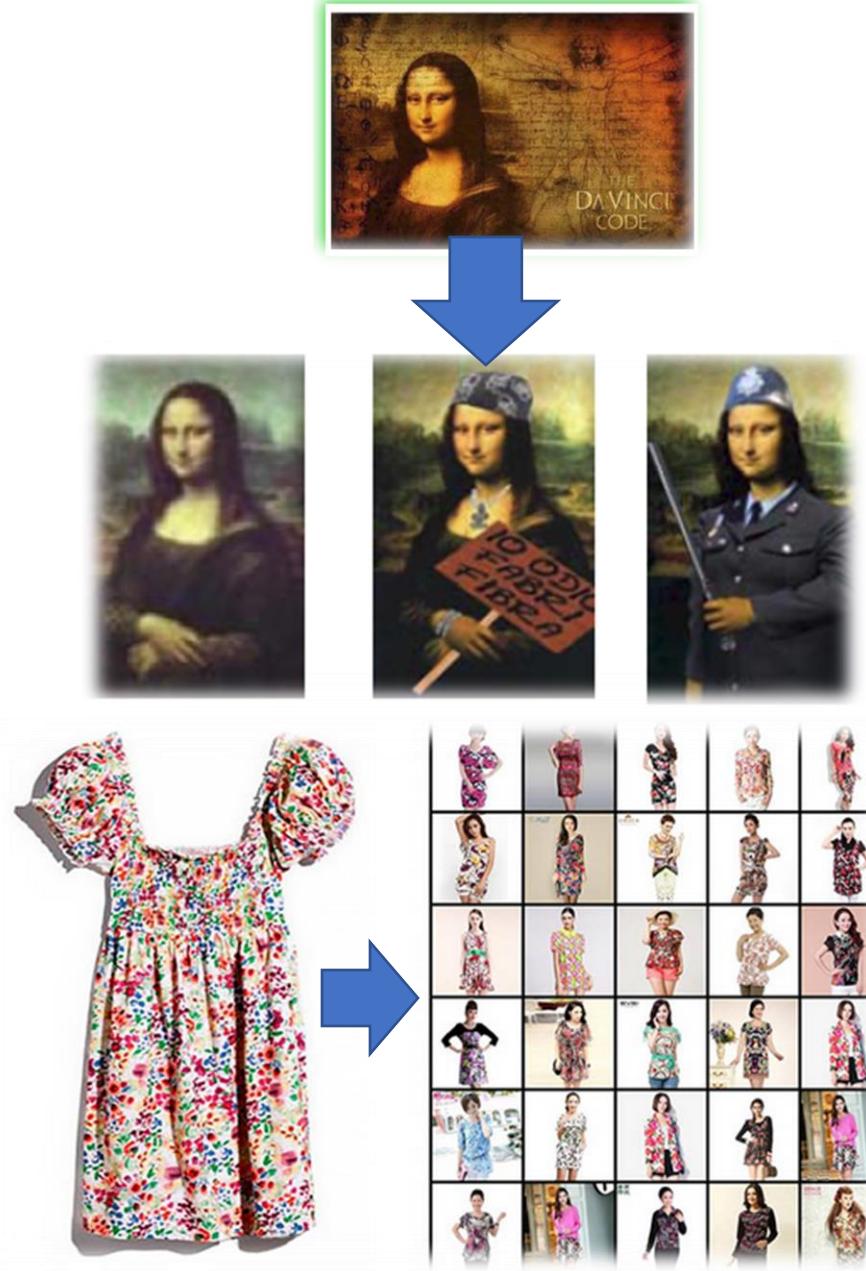
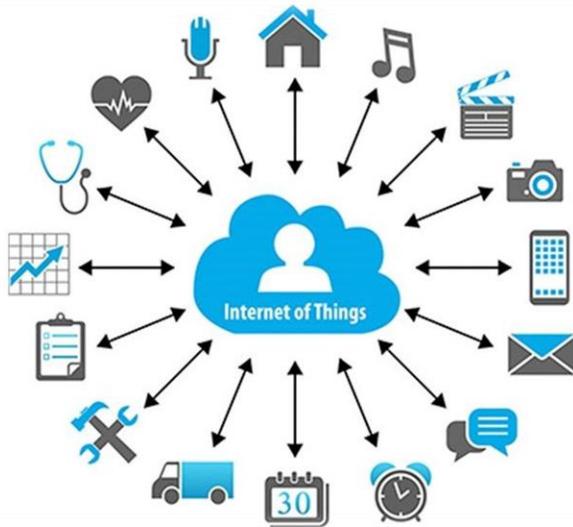
# Application

- ▶ Automotive safety and automatic car driving
- ▶ Robotics and Human-computer interaction



# Application

- ▶ Automotive safety and automatic car driving
- ▶ Robotics and Human-computer interaction
- ▶ Internet of Things



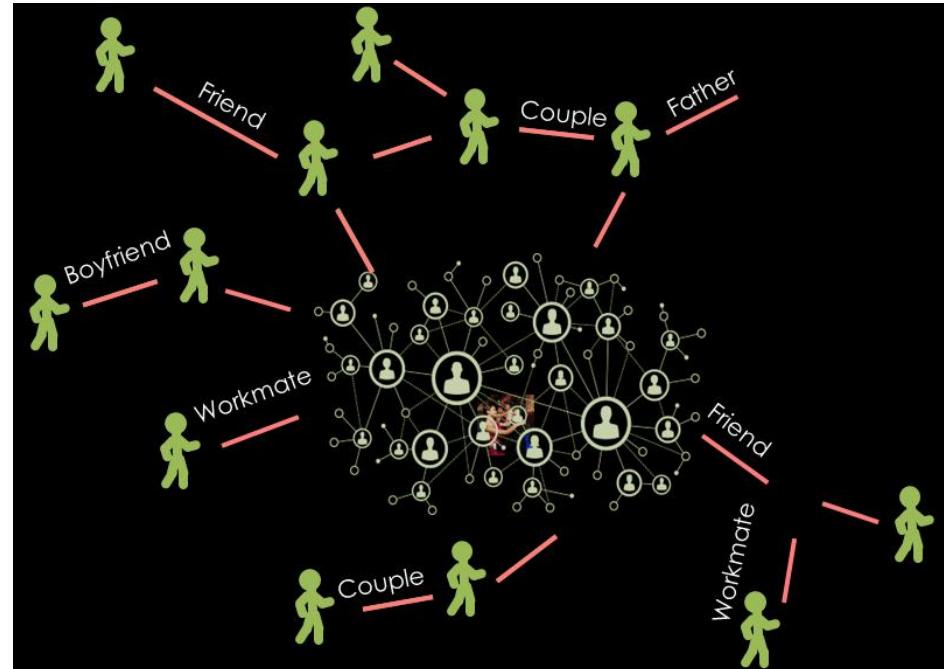
# Application

- ▶ Automotive safety and automatic car driving
- ▶ Robotics and Human-computer interaction
- ▶ Internet of Things
- ▶ Public safety and smart city



# Application

- ▶ Automotive safety and automatic car driving
- ▶ Robotics and Human-computer interaction
- ▶ Internet of Things
- ▶ Public safety and smart city
- ▶ Social network



Family



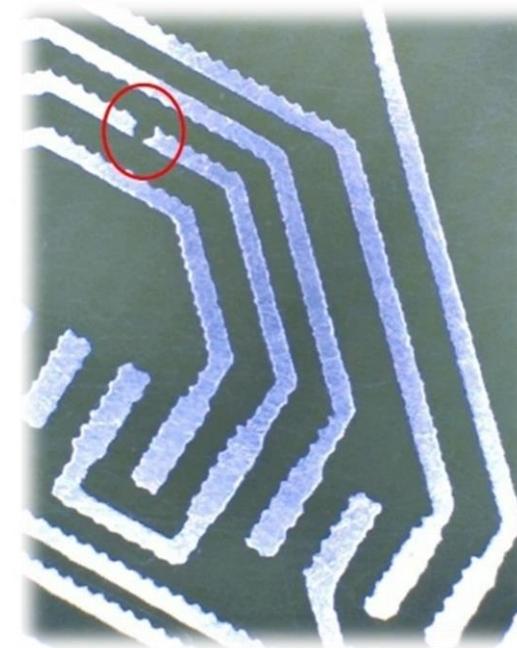
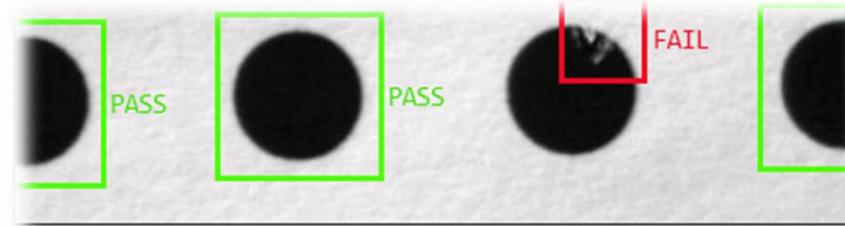
Workmate



Father

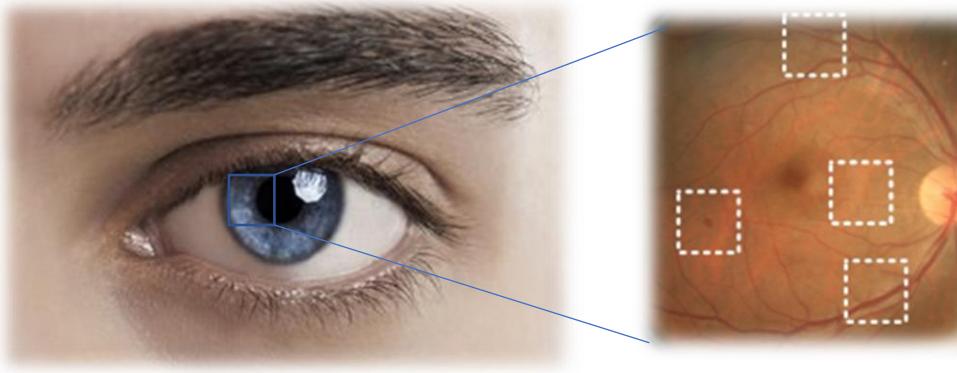
# Application

- ▶ Automotive safety and automatic car driving
- ▶ Robotics and Human-computer interaction
- ▶ Internet of Things
- ▶ Public safety and smart city
- ▶ Social network
- ▶ Industrial production

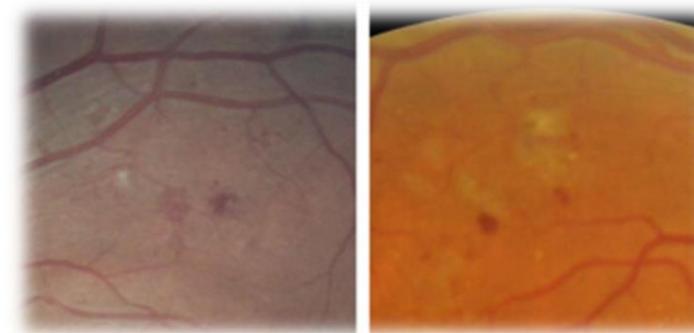


# Application

- ▶ Automotive safety and automatic car driving
- ▶ Robotics and Human-computer interaction
- ▶ Internet of Things
- ▶ Public safety and smart city
- ▶ Social network
- ▶ Industrial production
- ▶ Bio-medical imaging



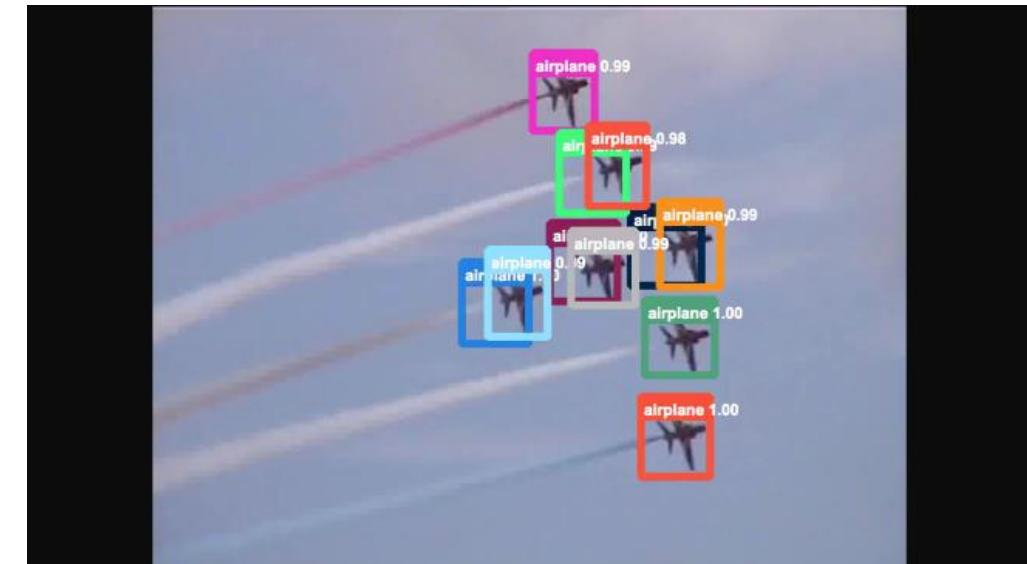
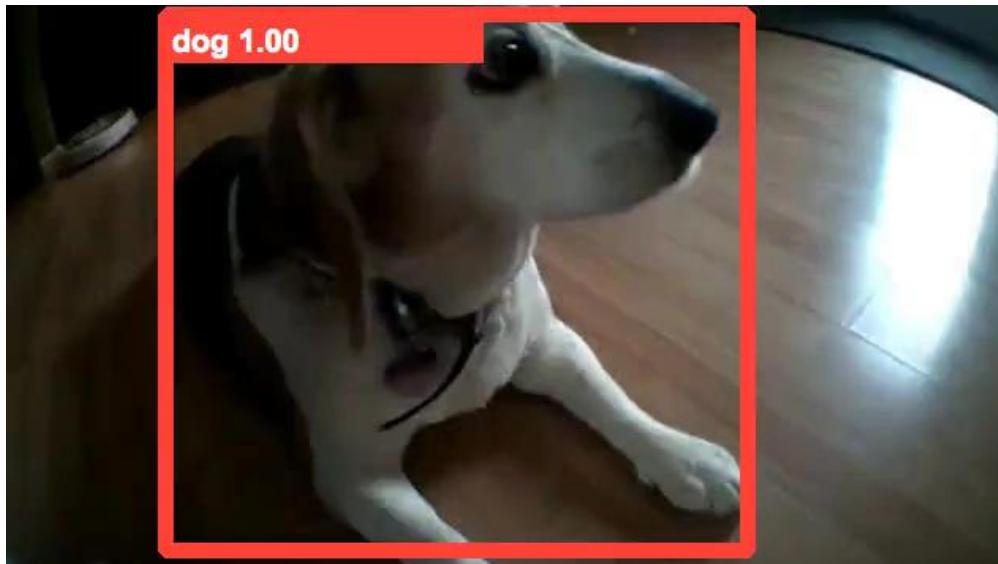
Microaneurysms



Blot hemorrhages

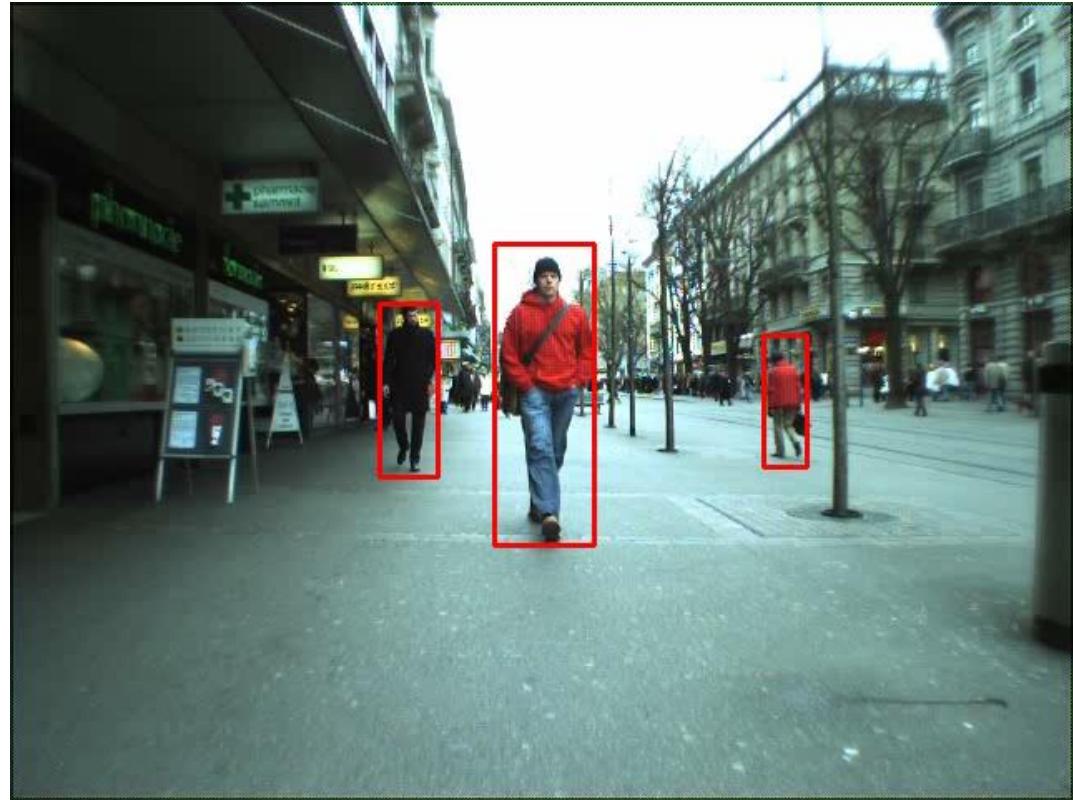
# Our team at ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

	ImageNet 2014	ImageNet 2015	ImageNet 2016	COCO 2018
Object detection	2nd (Google 1st)		1 <sup>st</sup>	1 <sup>st</sup>
Video object detection/tracking		1 <sup>st</sup>	1 <sup>st</sup>	



# We are the pioneer in deep learning

- We are among the first developing deep models for
  - pedestrian detection [1]



[1] Wanli Ouyang, and X. Wang. "A discriminative deep model for pedestrian detection with occlusion handling." In Proc. CVPR, 2012.

# We are the pioneer in deep learning

- We are among the first developing deep models for
  - pedestrian detection [1]
  - human pose estimation [2]



[1] Wanli Ouyang, and X. Wang. "A discriminative deep model for pedestrian detection with occlusion handling." In Proc. CVPR, 2012.

[2] Wanli Ouyang, Xiao Chu, Xiaogang Wang, "Multi-source Deep Learning for Human Pose Estimation", In Proc. CVPR 2014.

# We are the pioneer in deep learning

- We are among the first developing deep models for
  - pedestrian detection [1]
  - human pose estimation [2]
  - immediacy prediction [3]

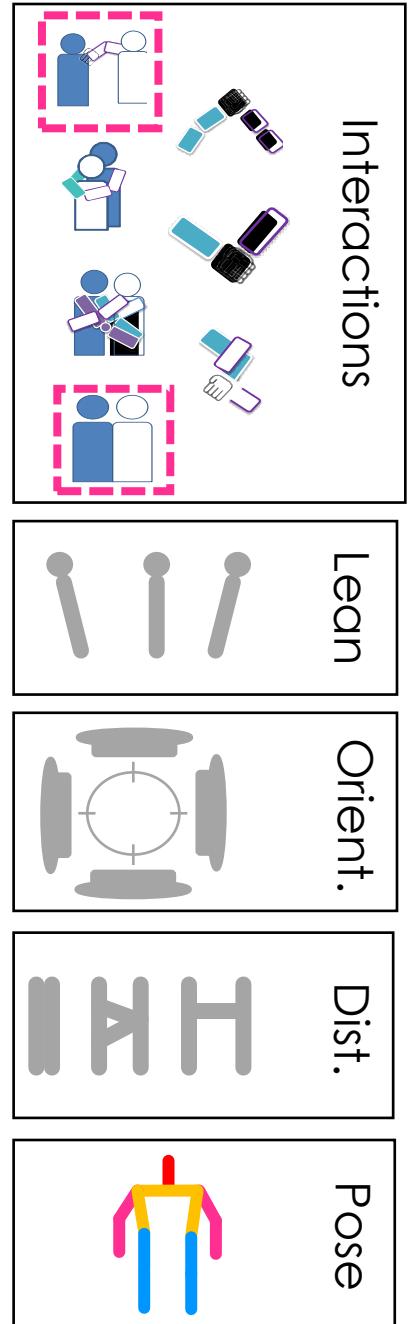


"shoulder to shoulder" with "arm over shoulder"

[1] Wanli Ouyang, and X. Wang. "A discriminative deep model for pedestrian detection with occlusion handling." In Proc. CVPR, 2012.

[2] Wanli Ouyang, Xiao Chu, Xiaogang Wang, "Multi-source Deep Learning for Human Pose Estimation", In Proc. CVPR 2014.

[3] Xiao Chu, Wanli Ouyang, Wei yang, Xiaogang Wang, "Multi-task Recurrent Neural Network for Immediacy Prediction", In Proc. ICCV 2015.



# We are the pioneer in deep learning

- We are among the first developing deep models for
  - pedestrian detection [1]
  - human pose estimation [2]
  - immediacy prediction [3]

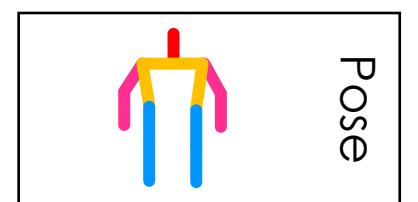
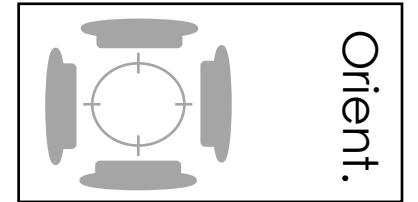
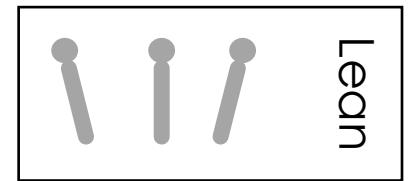
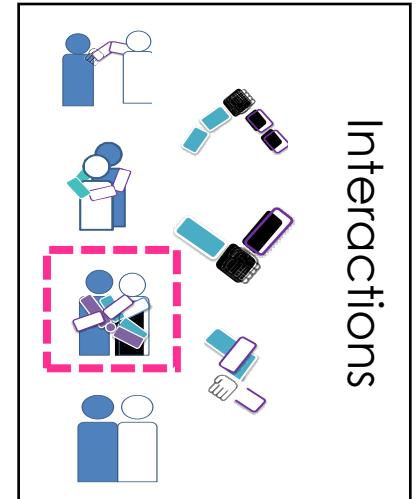


"Hug" at "close distance"

[1] Wanli Ouyang, and X. Wang. "A discriminative deep model for pedestrian detection with occlusion handling." In Proc. CVPR, 2012.

[2] Wanli Ouyang, Xiao Chu, Xiaogang Wang, "Multi-source Deep Learning for Human Pose Estimation", In Proc. CVPR 2014.

[3] Xiao Chu, Wanli Ouyang, Wei yang, Xiaogang Wang, "Multi-task Recurrent Neural Network for Immediacy Prediction", In Proc. ICCV 2015.



# We are the pioneer in deep learning

- We are among the first developing deep models for
  - pedestrian detection [1]
  - human pose estimation [2]
  - immediacy prediction [3]

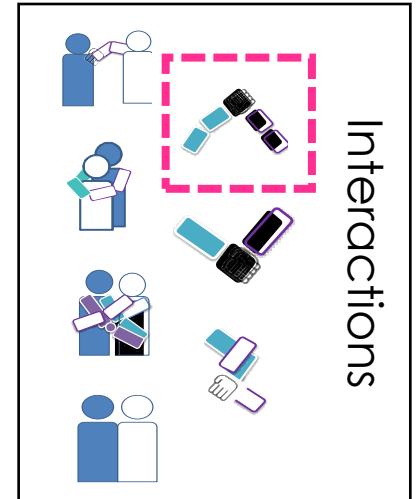


"High five" in a "longer distance"

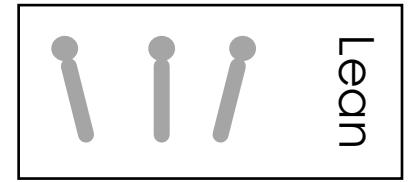
[1] Wanli Ouyang, and X. Wang. "A discriminative deep model for pedestrian detection with occlusion handling." In Proc. CVPR, 2012.

[2] Wanli Ouyang, Xiao Chu, Xiaogang Wang, "Multi-source Deep Learning for Human Pose Estimation", In Proc. CVPR 2014.

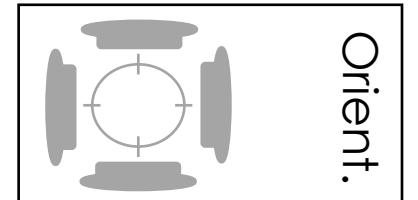
[3] Xiao Chu, Wanli Ouyang, Wei yang, Xiaogang Wang, "Multi-task Recurrent Neural Network for Immediacy Prediction", In Proc. ICCV 2015.



Interactions



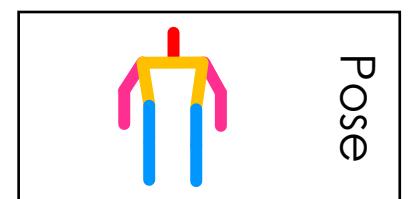
Lean



Orient.



Dist.



Pose

# We are the pioneer in deep learning

- We are among the first developing deep models for
  - pedestrian detection [1]
  - human pose estimation [2]
  - immediacy prediction [3]



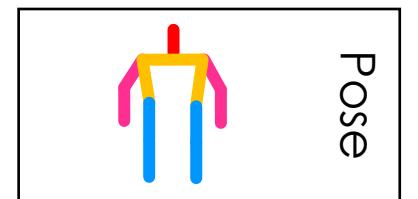
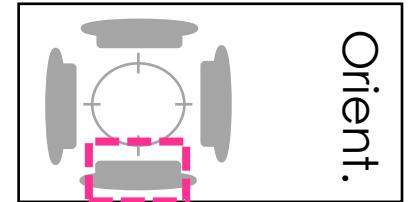
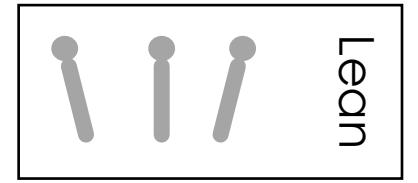
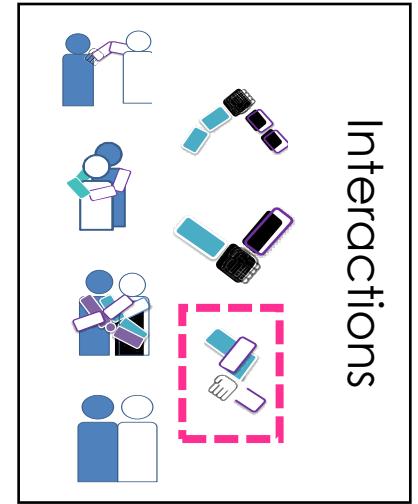
“Arm in arm” and “standing side by side”

[1] Wanli Ouyang, and X. Wang. "A discriminative deep model for pedestrian detection with occlusion handling." In Proc. CVPR, 2012.

[2] Wanli Ouyang, Xiao Chu, Xiaogang Wang, "Multi-source Deep Learning for Human Pose Estimation", In Proc. CVPR 2014.

[3] Xiao Chu, Wanli Ouyang, Wei yang, Xiaogang Wang, "Multi-task Recurrent Neural Network for Immediacy Prediction", In Proc. ICCV 2015.

[4] Rui Zhao, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang, "Saliency Detection by Multi-context Deep Learning", In Proc. CVPR, 2015.



# We are the pioneer in deep learning

- We are among the first developing deep models for
  - pedestrian detection [1]
  - human pose estimation [2]
  - immediacy prediction [3]
  - salient object detection [4]



[1] Wanli Ouyang, and X. Wang. "A discriminative deep model for pedestrian detection with occlusion handling." In Proc. CVPR, 2012.

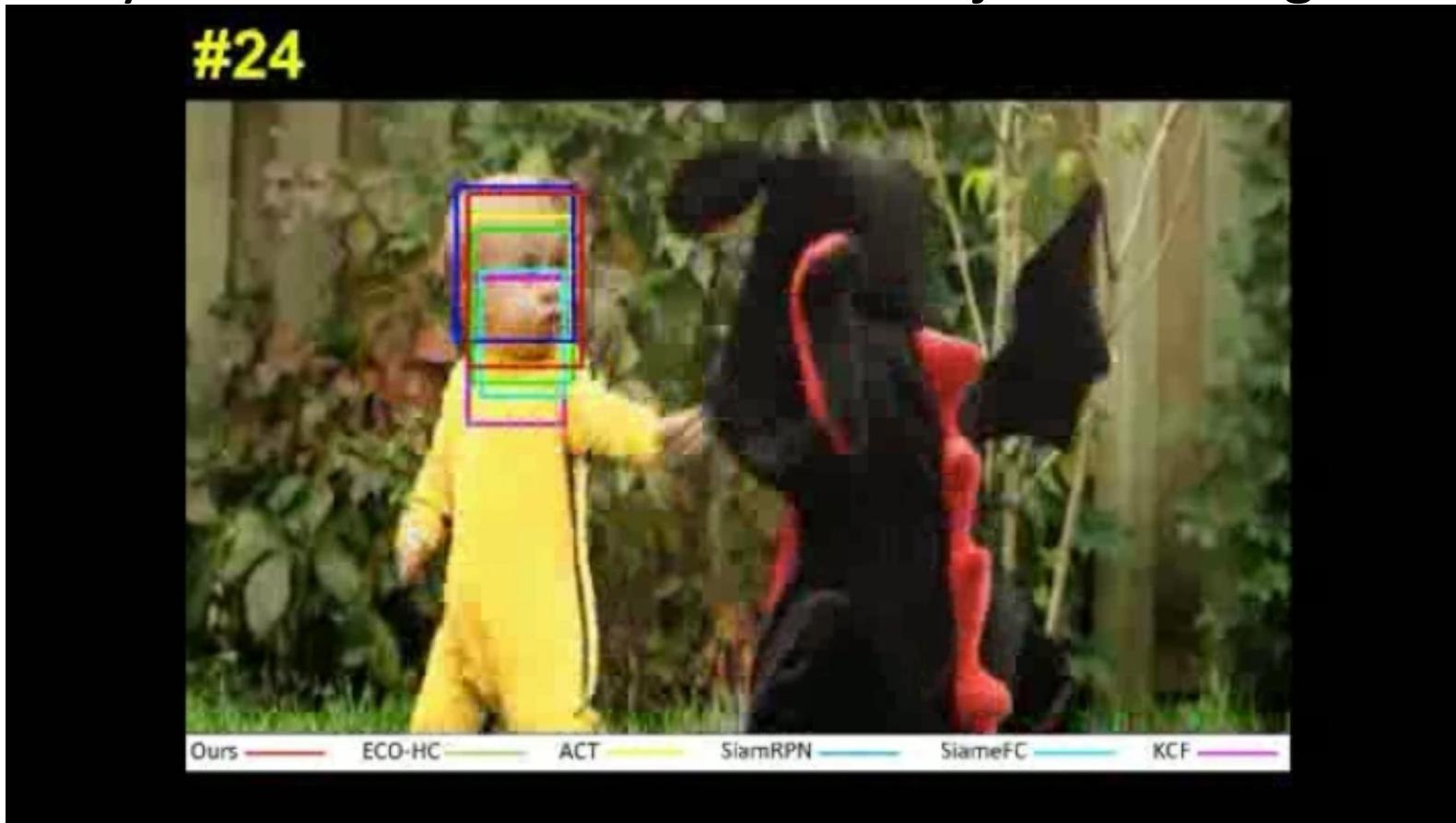
[2] Wanli Ouyang, Xiao Chu, Xiaogang Wang, "Multi-source Deep Learning for Human Pose Estimation", In Proc. CVPR 2014.

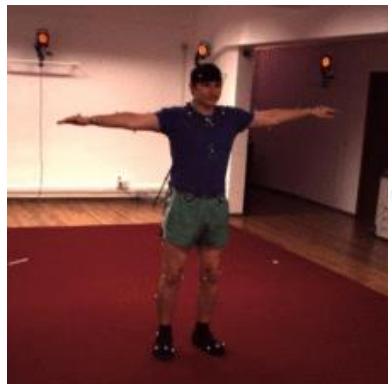
[3] Xiao Chu, Wanli Ouyang, Wei yang, Xiaogang Wang, "Multi-task Recurrent Neural Network for Immediacy Prediction", In Proc. ICCV 2015.

[4] Rui Zhao, Wanli Ouyang, Hongsheng Li, and Xiaogang Wang, "Saliency Detection by Multi-context Deep Learning", In Proc. CVPR, 2015.

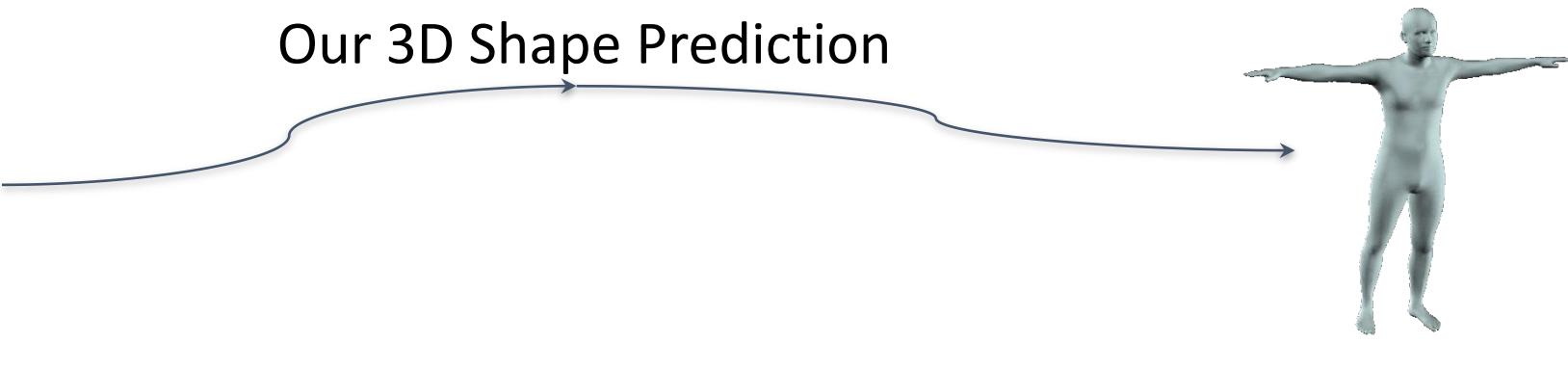
# We are the pioneer in deep learning

- The first fully convolutional network for object tracking





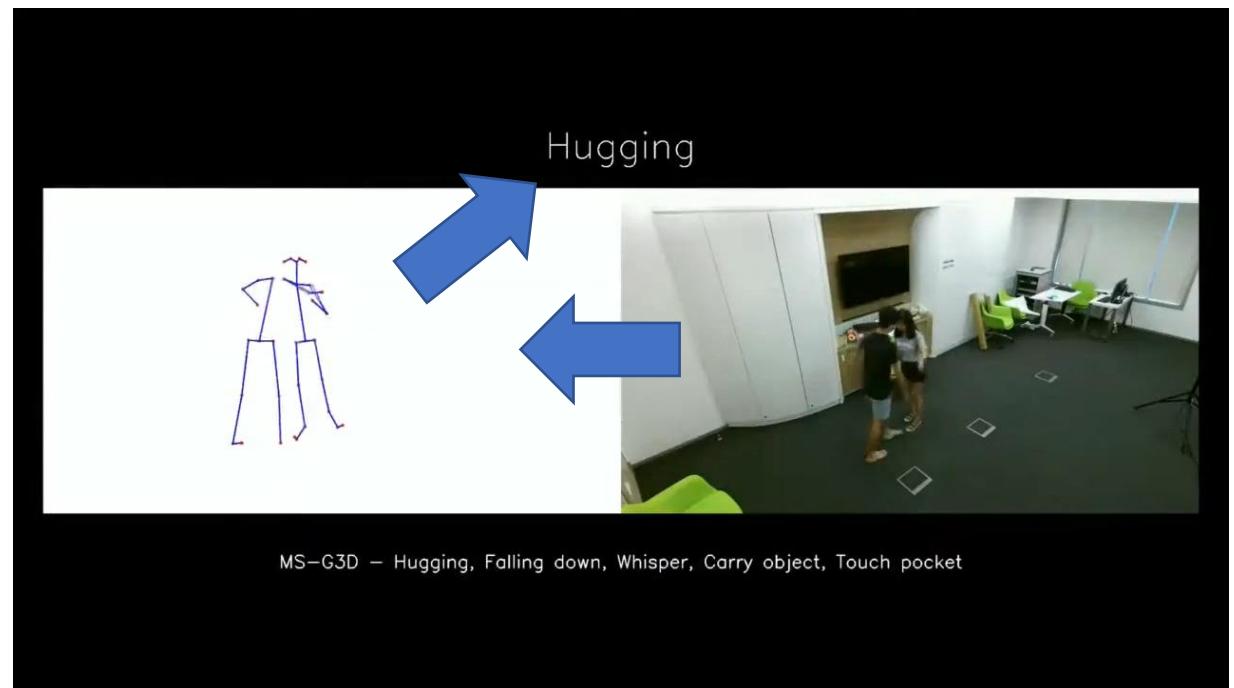
Input image



# Estimate Human Action Type

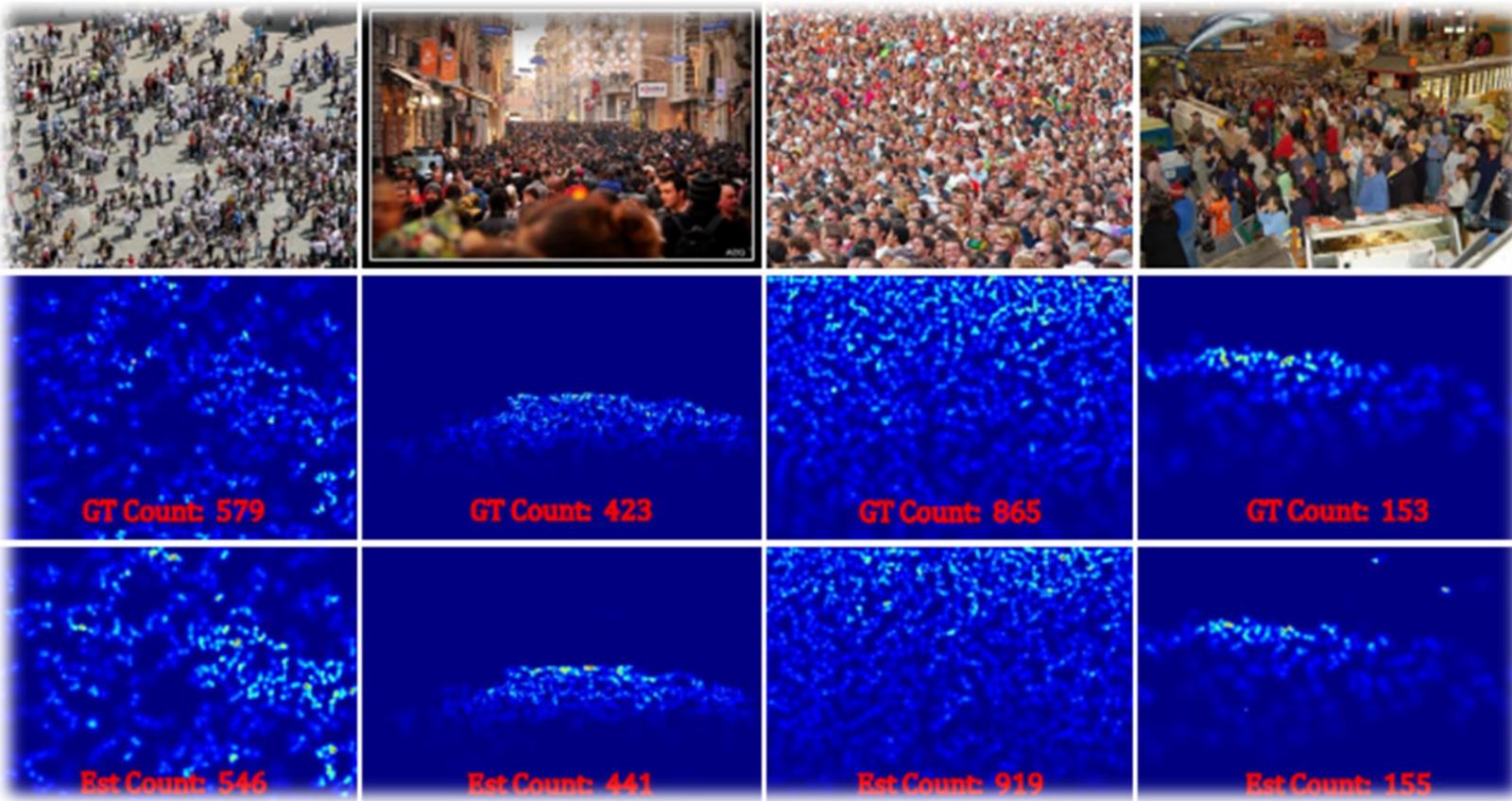


Shuyang Sun, Zhanghui Kuang, Lu Sheng, Wanli Ouyang, Wei Zhang, "Optical Flow Guided Feature: A Motion Representation for Video Action Recognition", Proc. CVPR, 2018



"Spatial-Temporal Graph Convolutions with Disentangled Multi-Scale Aggregators for Skeleton-Based Action Recognition", CVPR, 2020

# Crowd counting



Lingbo Liu, Zhilin Qiu , Guanbin Li, Shufan Liu, W. Ouyang, Liang Lin, "Crowd Counting with Deep Structured Scale Integration Network", Proc. ICCV, 2019

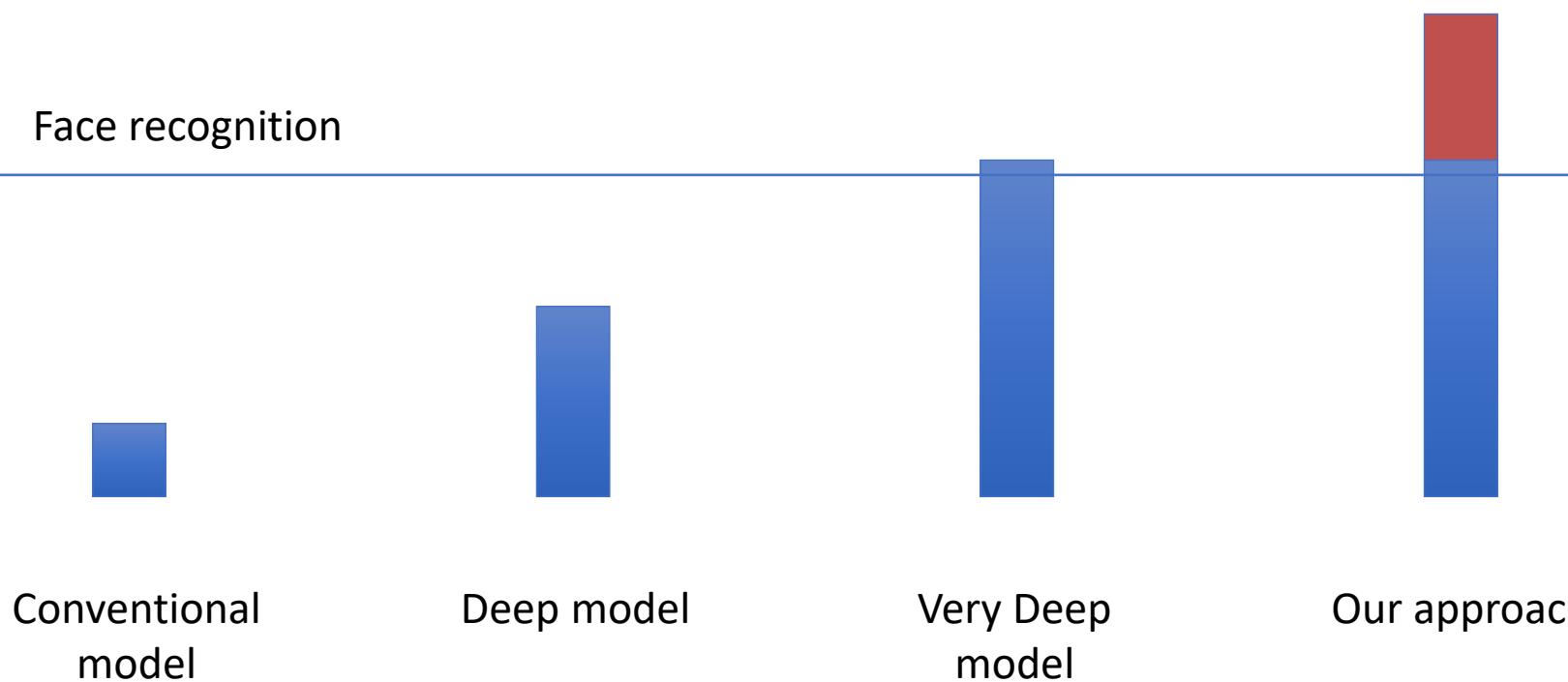
# Performance vs practical need

Many other applications

---

Face recognition

---



Conventional  
model

Deep model

Very Deep  
model

Our approach

# Outline

Introduction

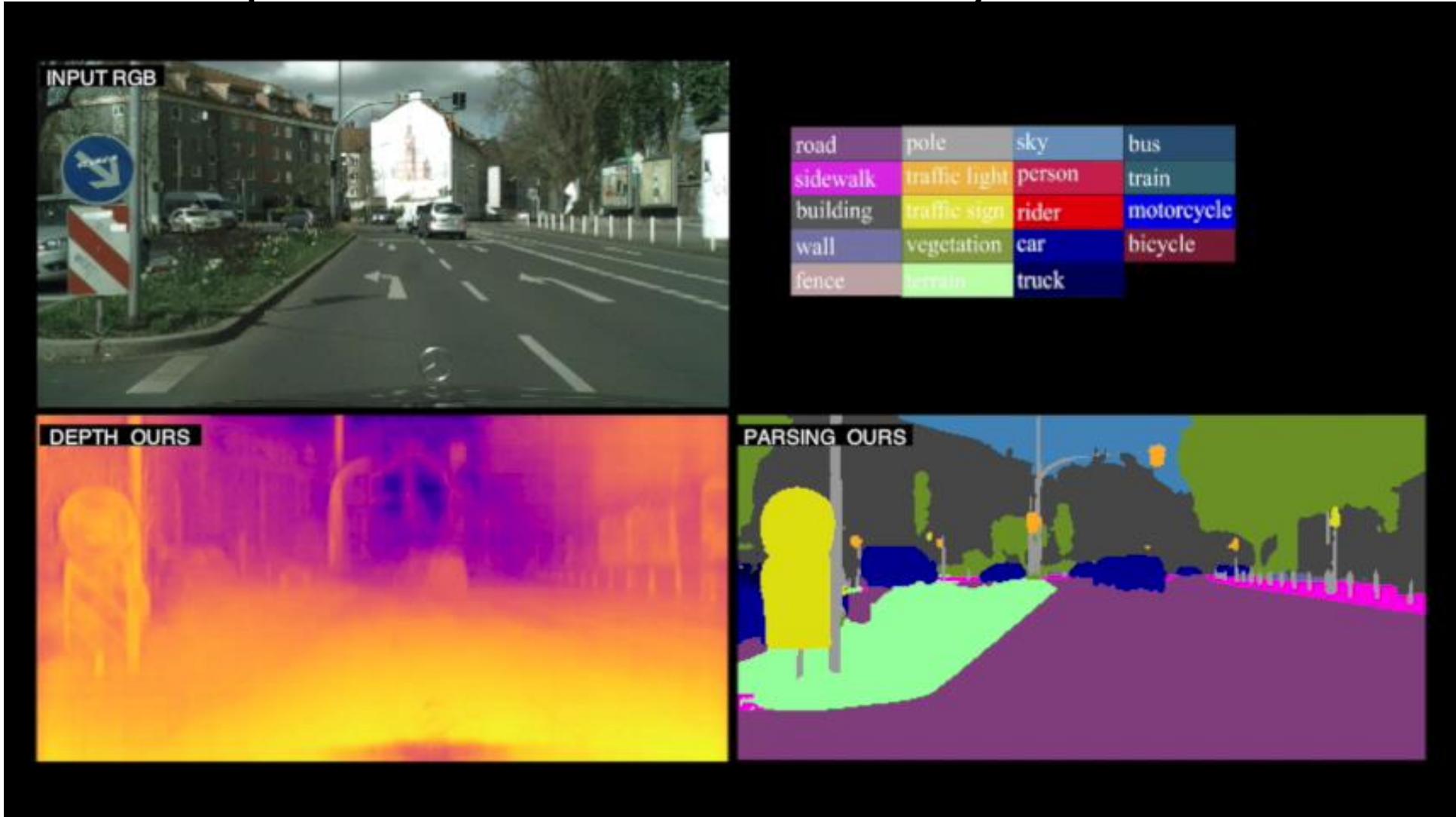
Our Research Works

Research topics

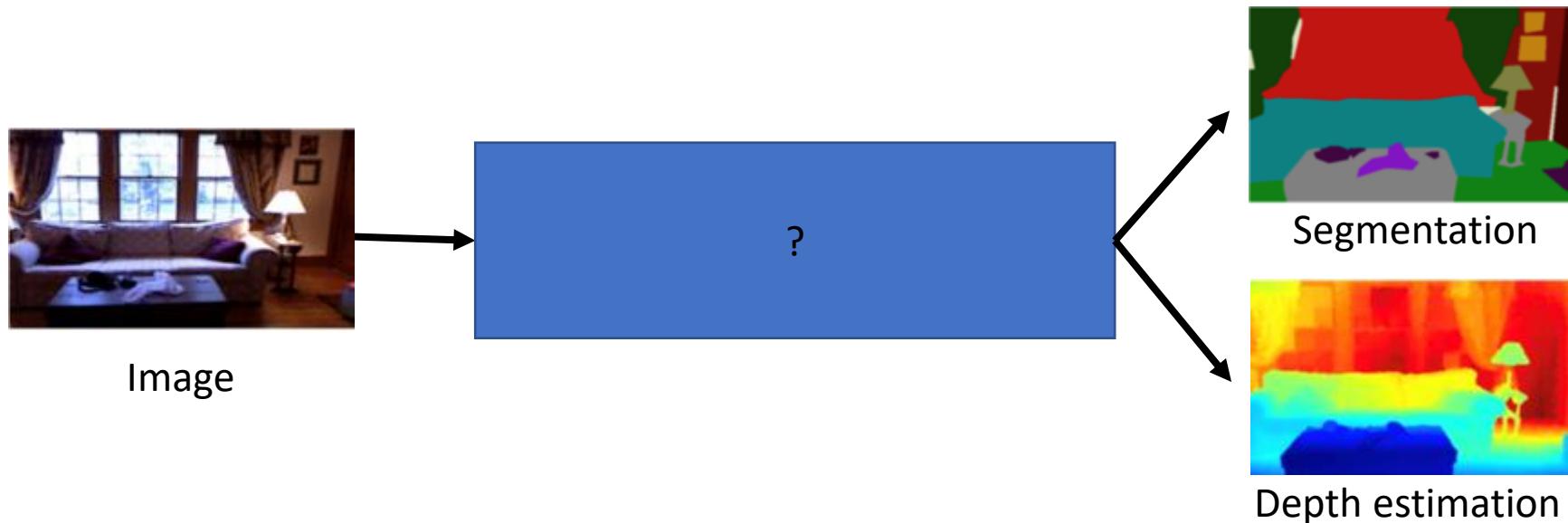
Research findings



# Estimate depth and semantic object class



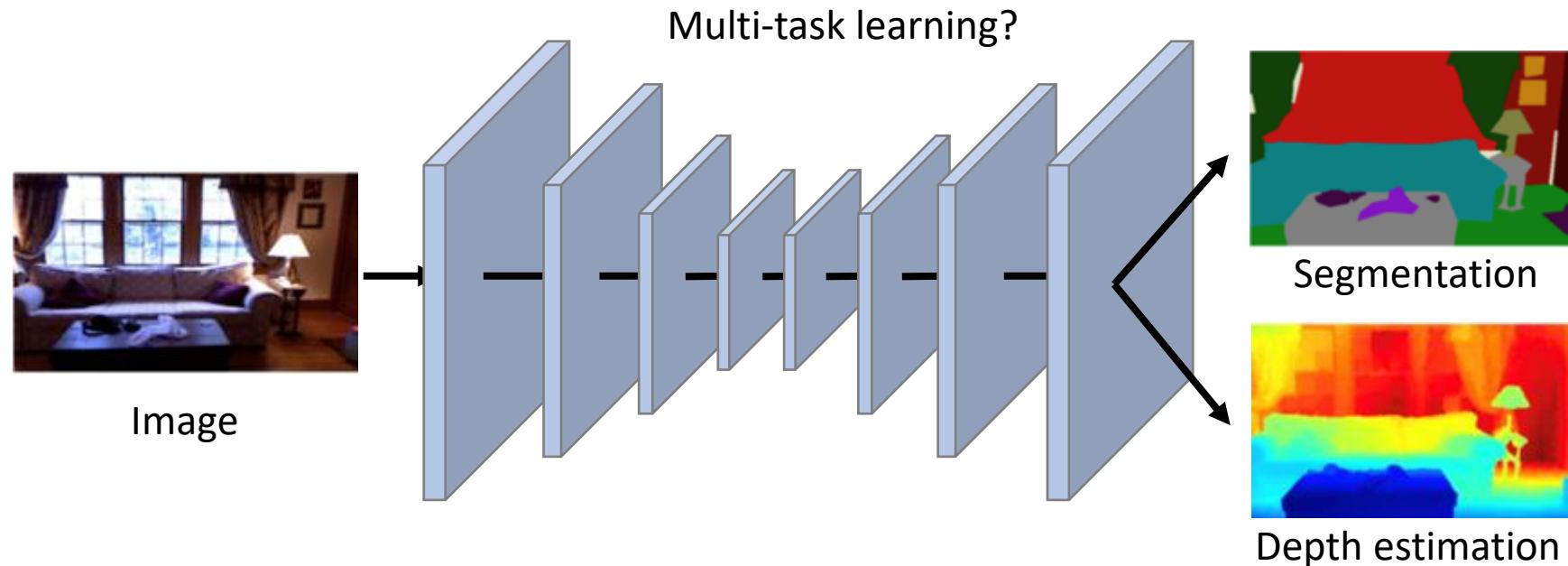
# From multi-task learning to multi-modal learning



# From multi-task learning to multi-modal learning

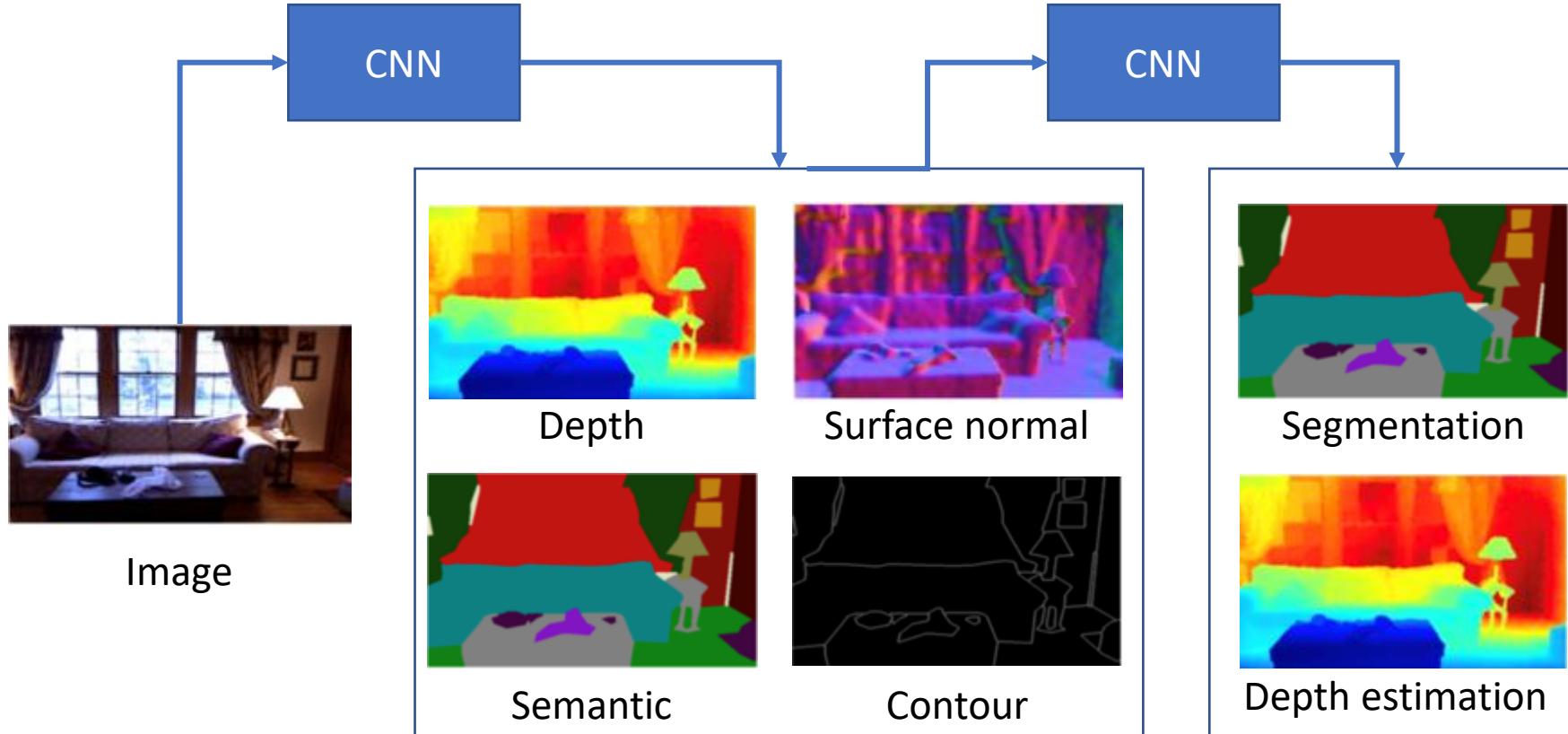
Multi-task learning: One AI approach, multiple tasks

- Directly optimizing multiple tasks given input training data does not guarantee consistent gain on all the tasks

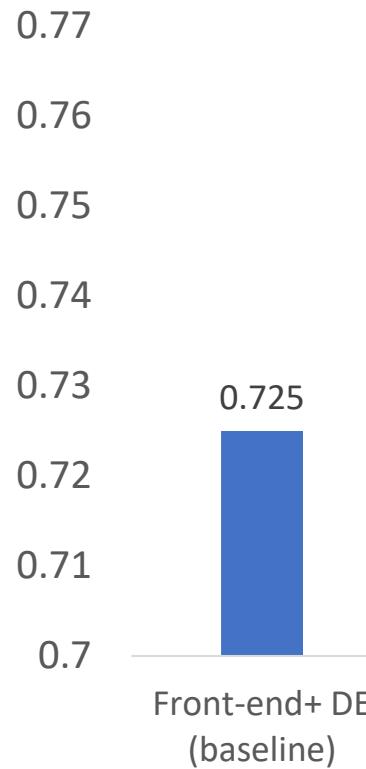


# From multi-task learning to multi-modal learning

Our novelty: Multi-modal input data improve training of AI algorithms



# Results for scene parsing on Cityscapes



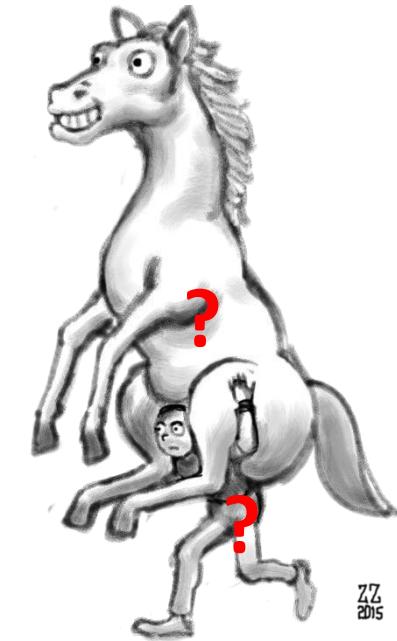
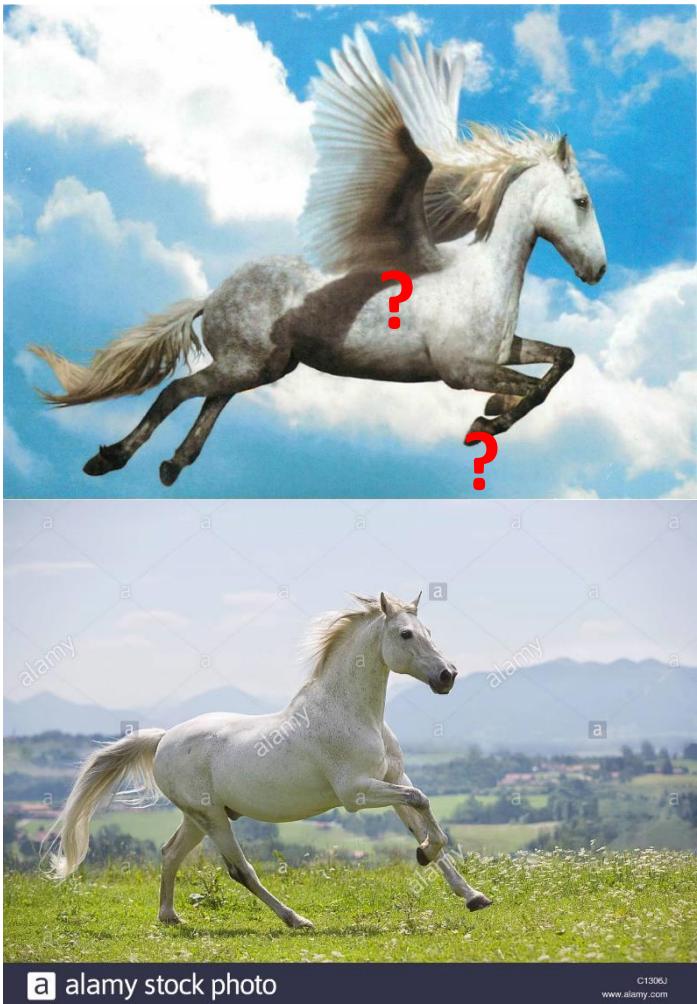
Multi-task learning results in decrease of IOU accuracy  
Our approach improves accuracy

# Structure Feature Learning

- Structure: the arrangement of and relations between the parts or elements of something complex<sup>1</sup>.
- Elements or parts are correlated.

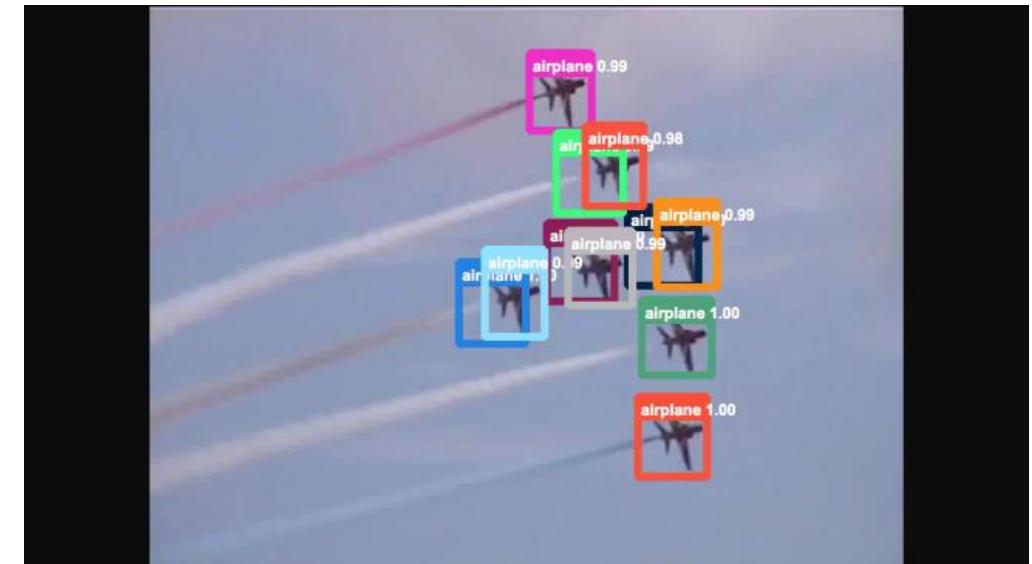
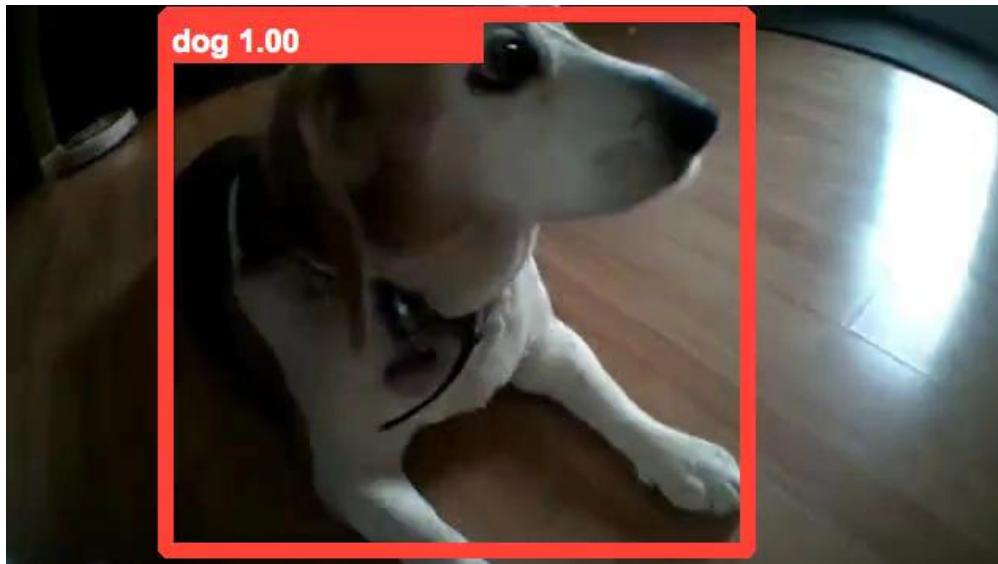
1. <https://en.wikipedia.org/wiki/Structure>

# Structure in data

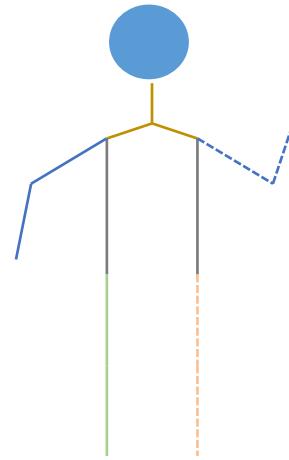
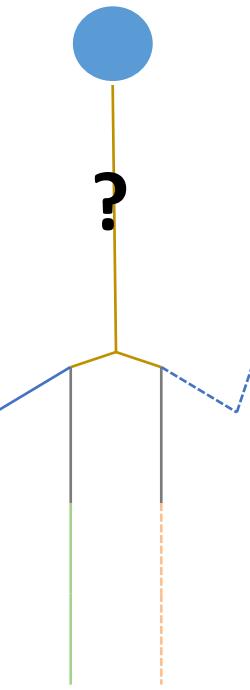
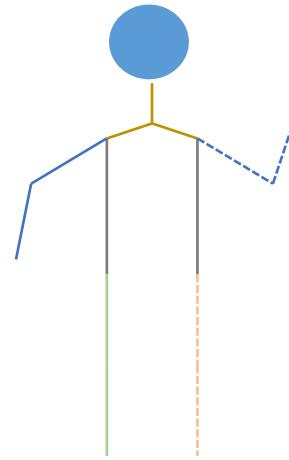
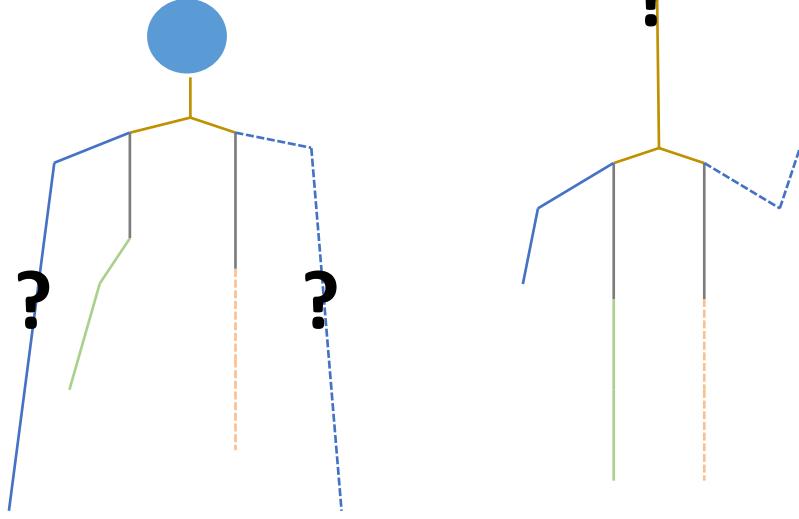


# Our team at ImageNet Large Scale Visual Recognition Challenge (ILSVRC)

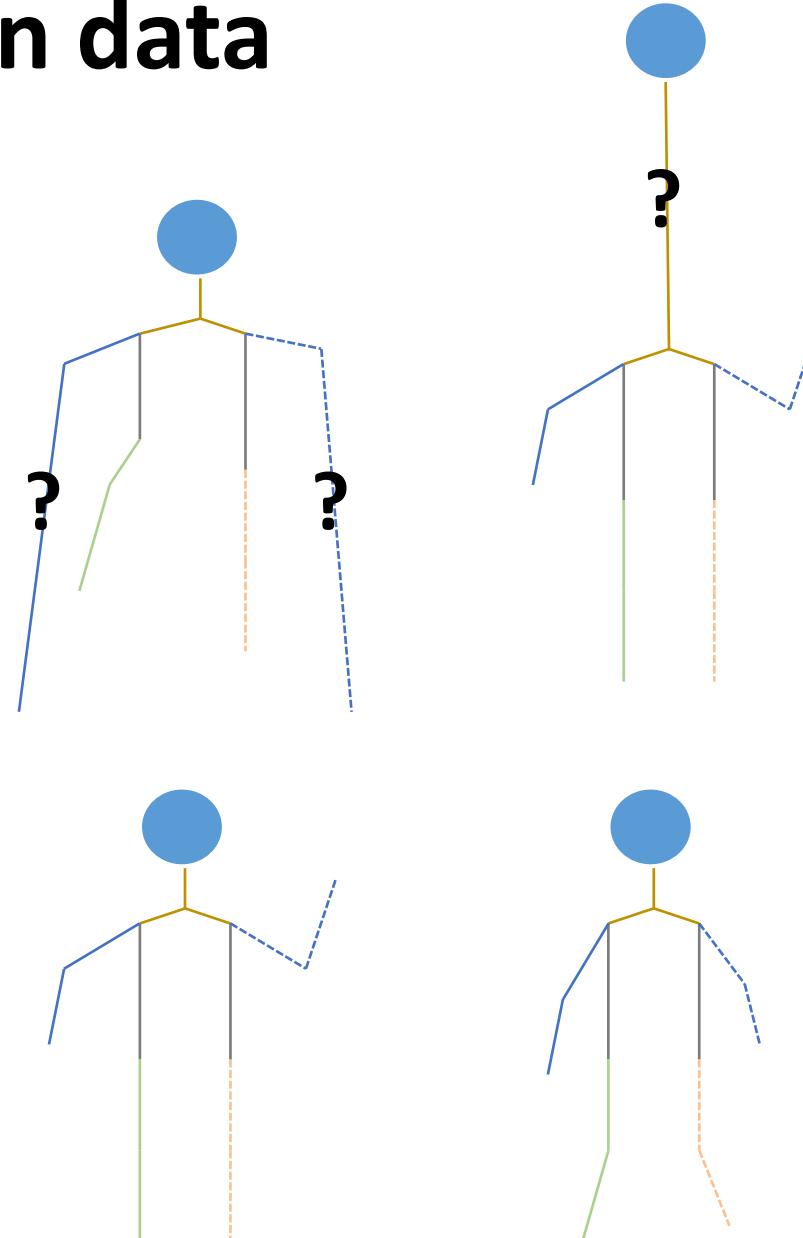
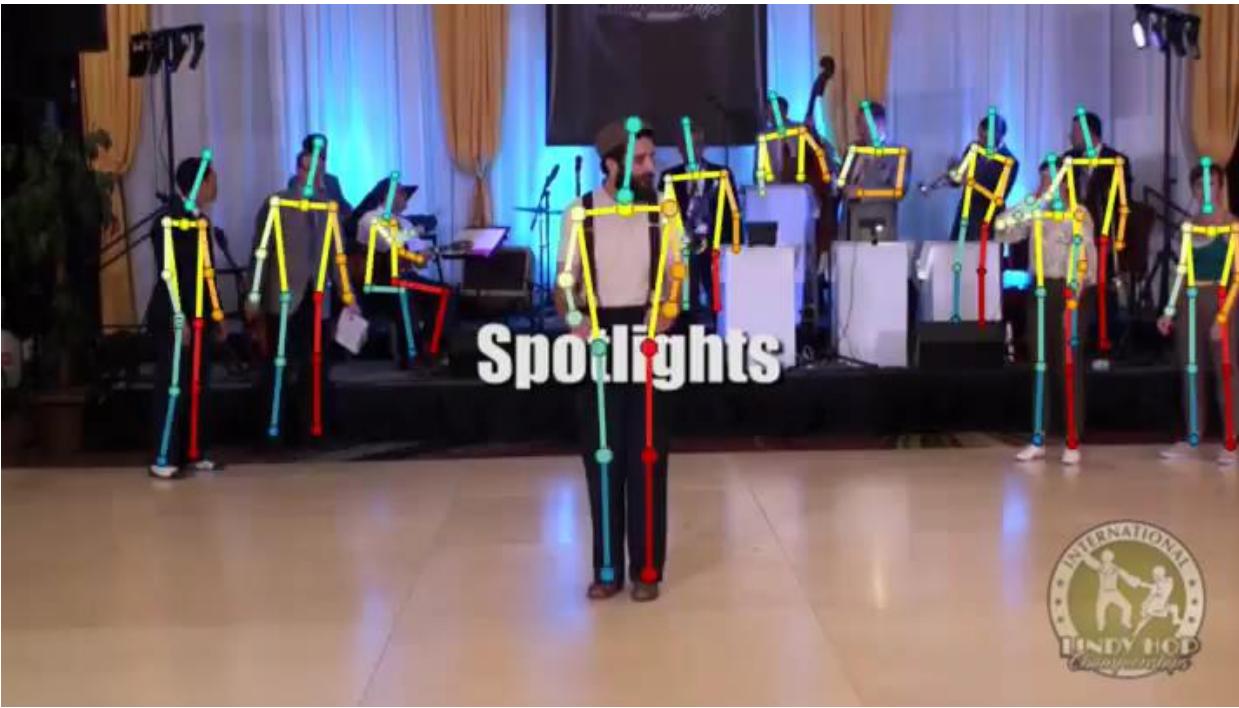
ImageNet 2016	
Object detection	1 <sup>st</sup>
Video object detection/tracking	1 <sup>st</sup>



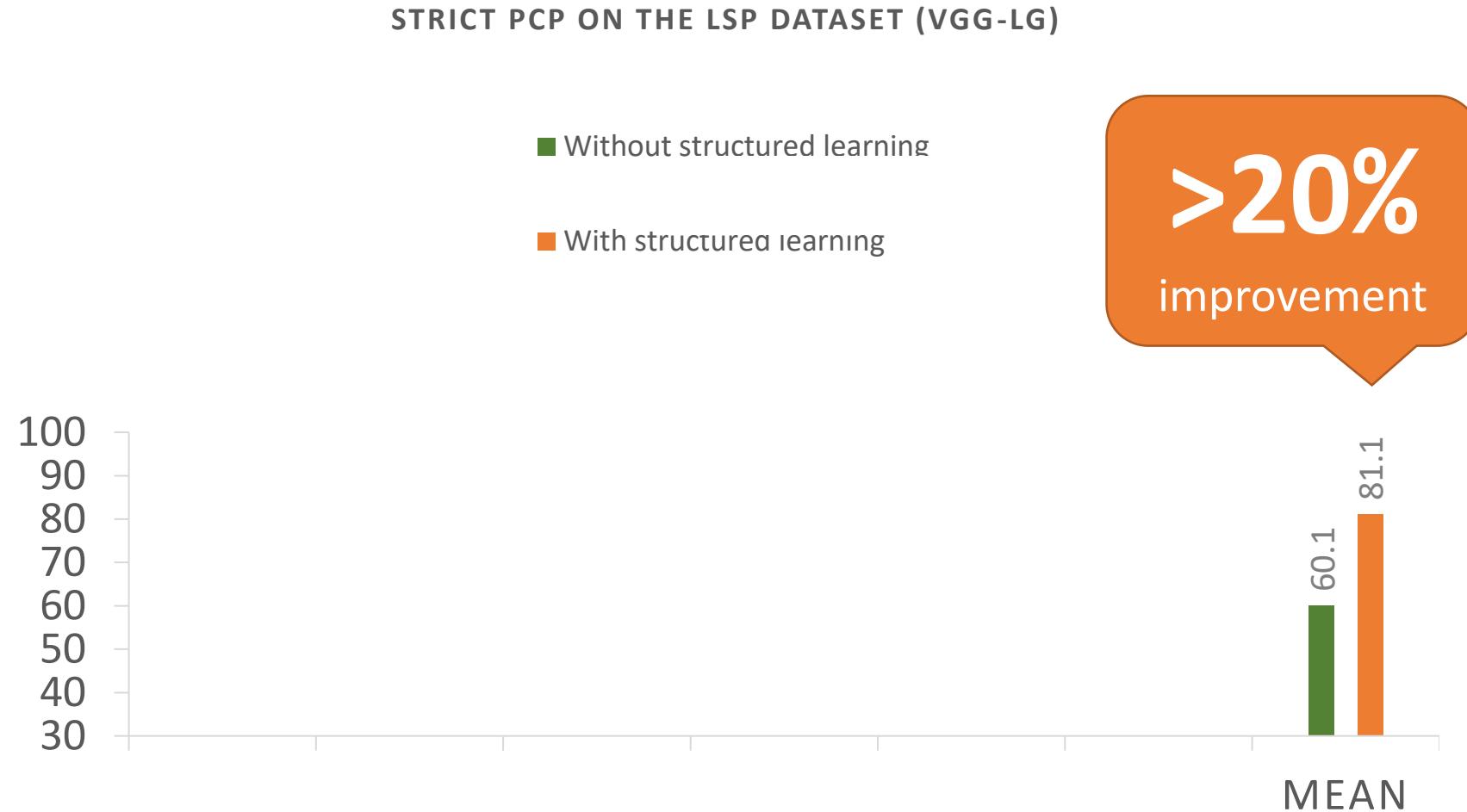
# Structure in data



# Structure in data



# Unary Term vs. Full Model



# Profound experience required

- We are experienced
  - Starting from 2011
  - \$4M medical research future fund (MRFF) of National health and medical research council (NHMRC)
  - Discovery fund (Australia Research Council)
  - \$1M direct industry fund
- Lower cost and faster for verify some new ideas for your company

# Outline

Introduction

Our Research Works

Research topics

Research findings

Discussion



# AI replace human?



<https://www.amazon.com/Terminator-Complete-Movie-Collection-Bonus/dp/B083ZK519Z>

# AI replace human?



<https://www.fwi.co.uk/farm-life/stay-safe-riding-horse-road>

# AI replace human?



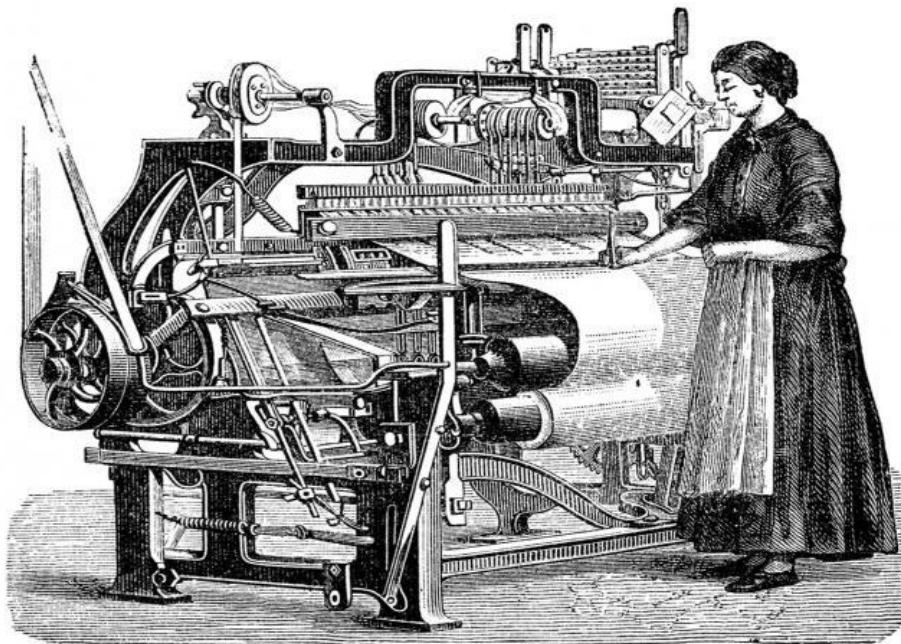
VS



<https://www.trendhunter.com/trends/old-man-lifts-fridge>

<https://2easyremovals.com.au/5-things-consider-moving-fridge/>

# AI replace human?



VS



<https://fashion-history.lovetoknow.com/fashion-clothing-industry/weaving-machinery>

<https://www.pinterest.com.au/pin/535083999468463019/>

# AI replace human?

- AI = New technology
- As other new technologies, AI is developed for benefiting us
  - Some jobs will disappear, e.g. drivers, more in [bit.ly/31efsye](https://bit.ly/31efsye)
  - Some other jobs will appear, e.g. AI business development manager, more in [bit.ly/3gqo6QI](https://bit.ly/3gqo6QI)
- AI can learn on
  - Recognizing objects
  - Understanding environments
  - Planning on specific tasks, such as driving, pick-and-place, playing games
- Currently, AI cannot replace human to
  - Enjoy life
  - Develop new technologies
  - Make critical decisions, e.g. to marry John or Smith?
  - Smoothly manage a team
  - Make this presentation

# Human ethics issues

- Privacy preserving sensors
- Privacy preserving algorithms
- Go through strict human ethics approval

# Outline

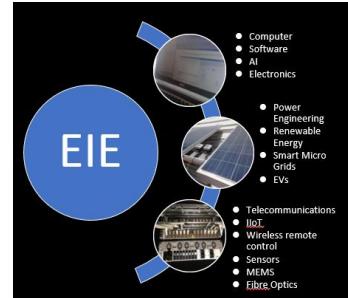


## Introduction

### Our Research Works

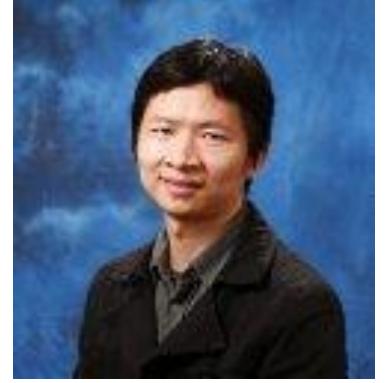


### My school



# Me

- Deputy head of school on Research
  - School of Electrical and Information Engineering

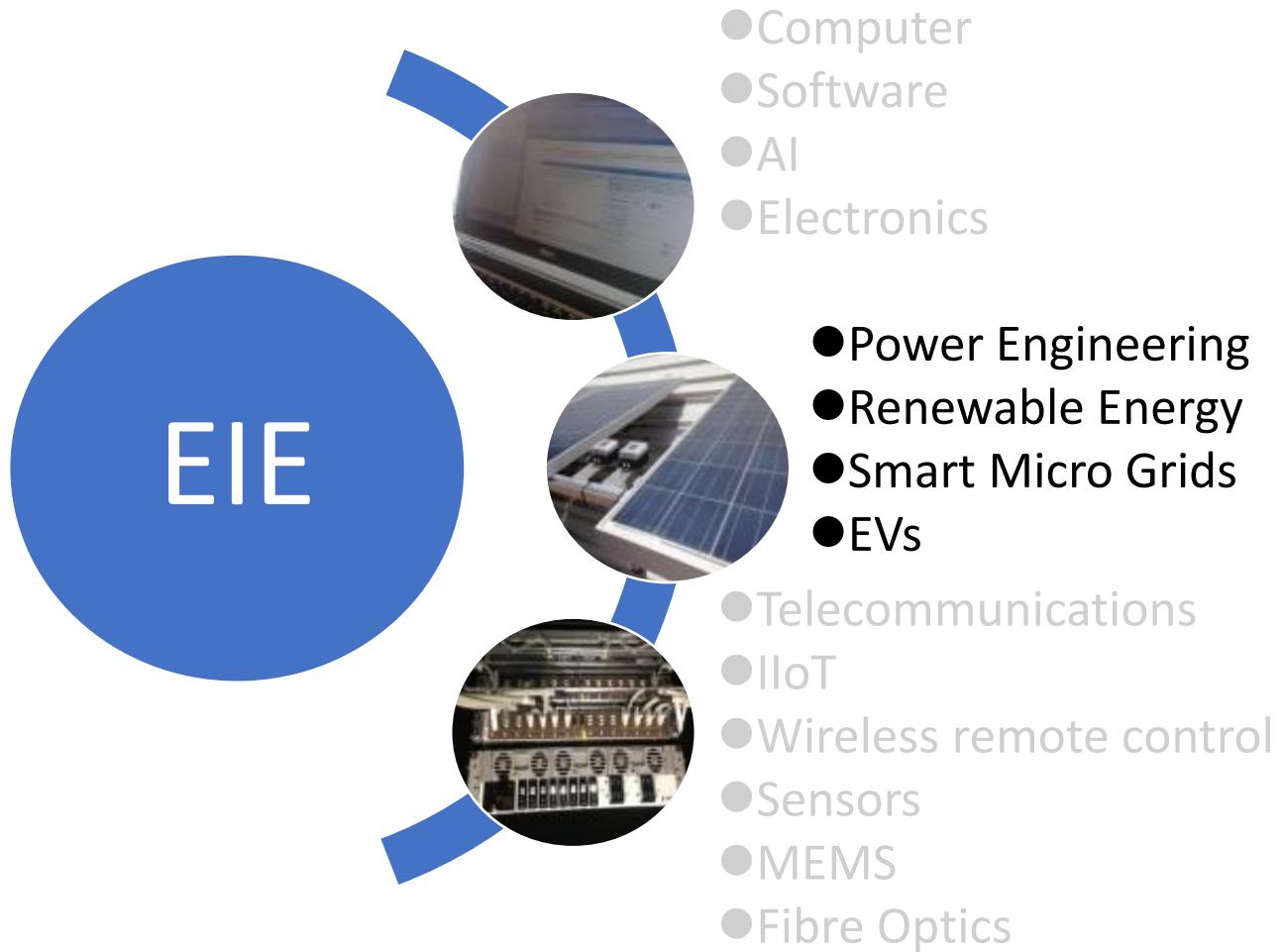


The University of Sydney

# What is Electrical and Information Engineering



# What is **Electrical** and Information Engineering

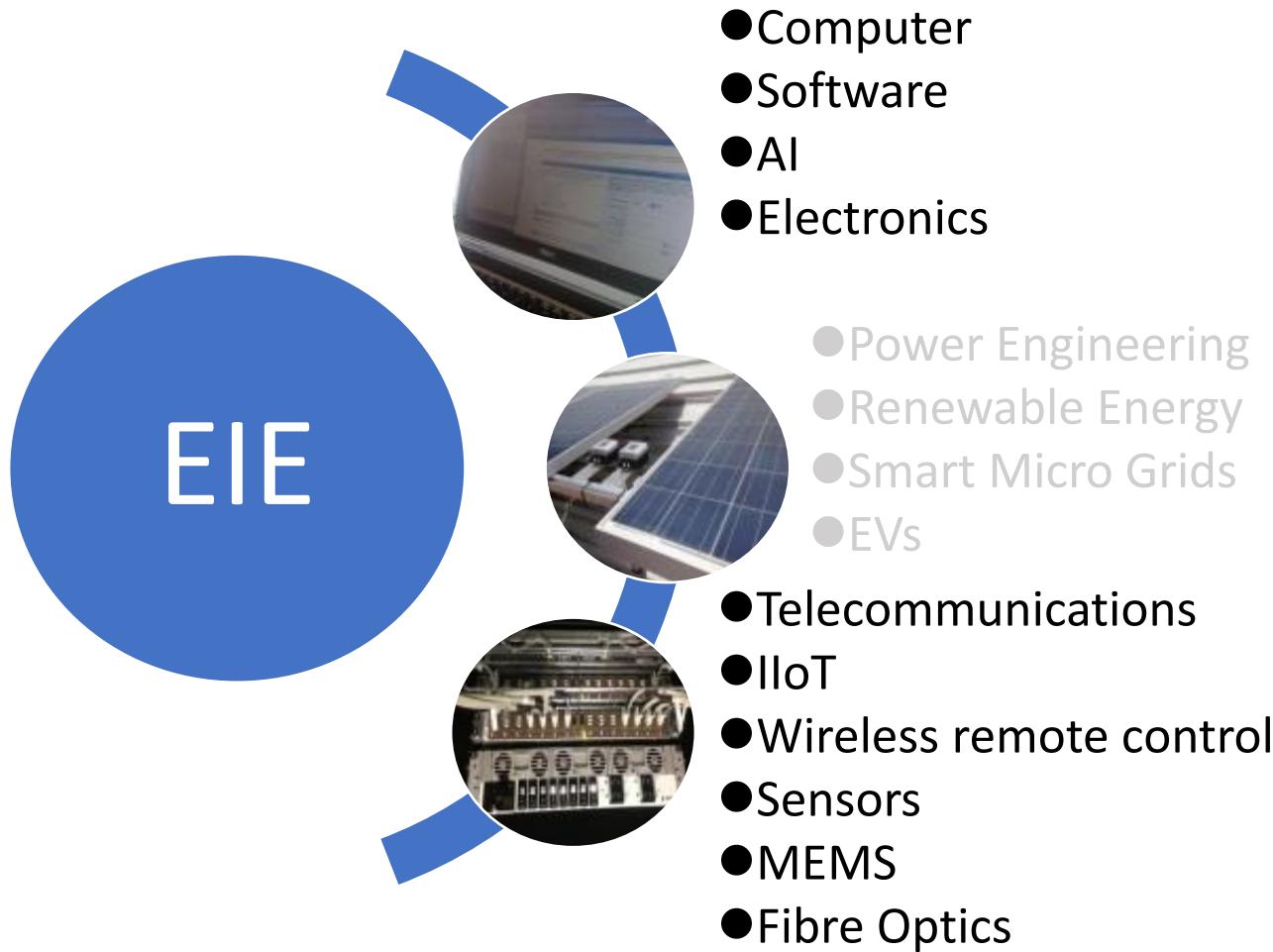


# Energy and Power engineering

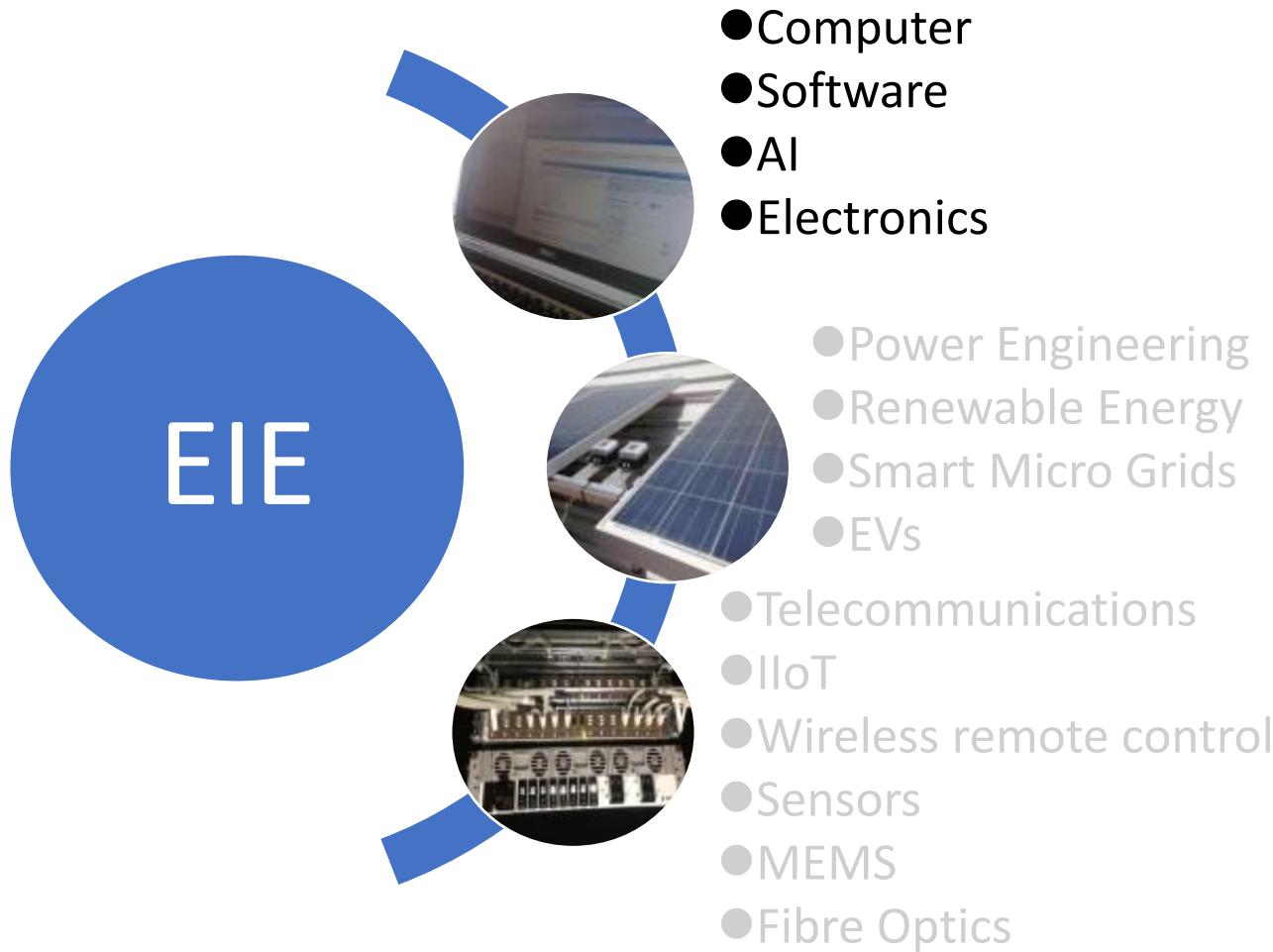
- Generation, transmission, and distribution of electrical energy
- Renewable energy sources
- High efficiency conversion, transmission, monitoring and control
- Power electronic devices vital for integration
- Advanced ICT (IoT in future) to coordinate
- Supported by the power industry – high demand from the energy sector



# What is Electrical and **Information** Engineering

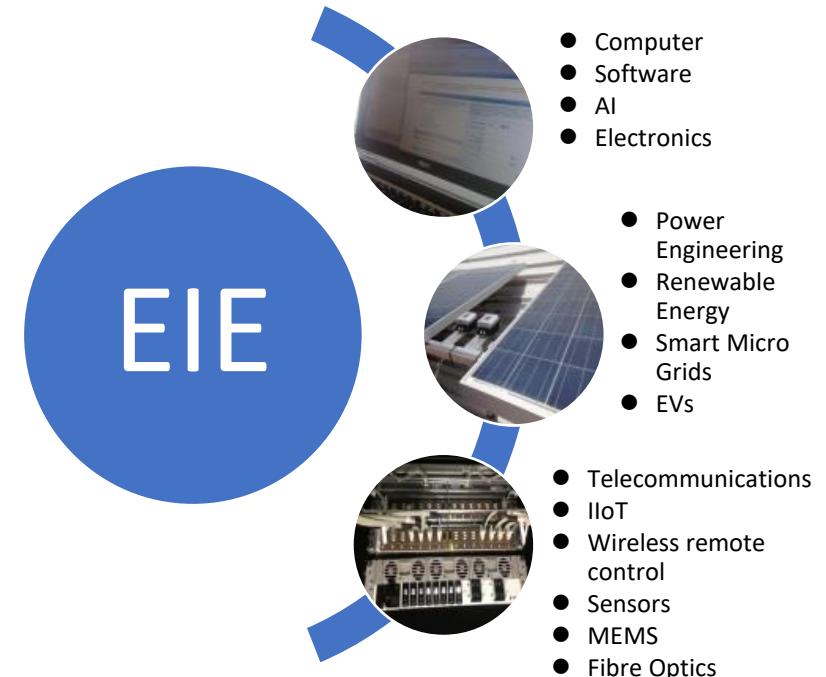


# What is Electrical and **Information** Engineering

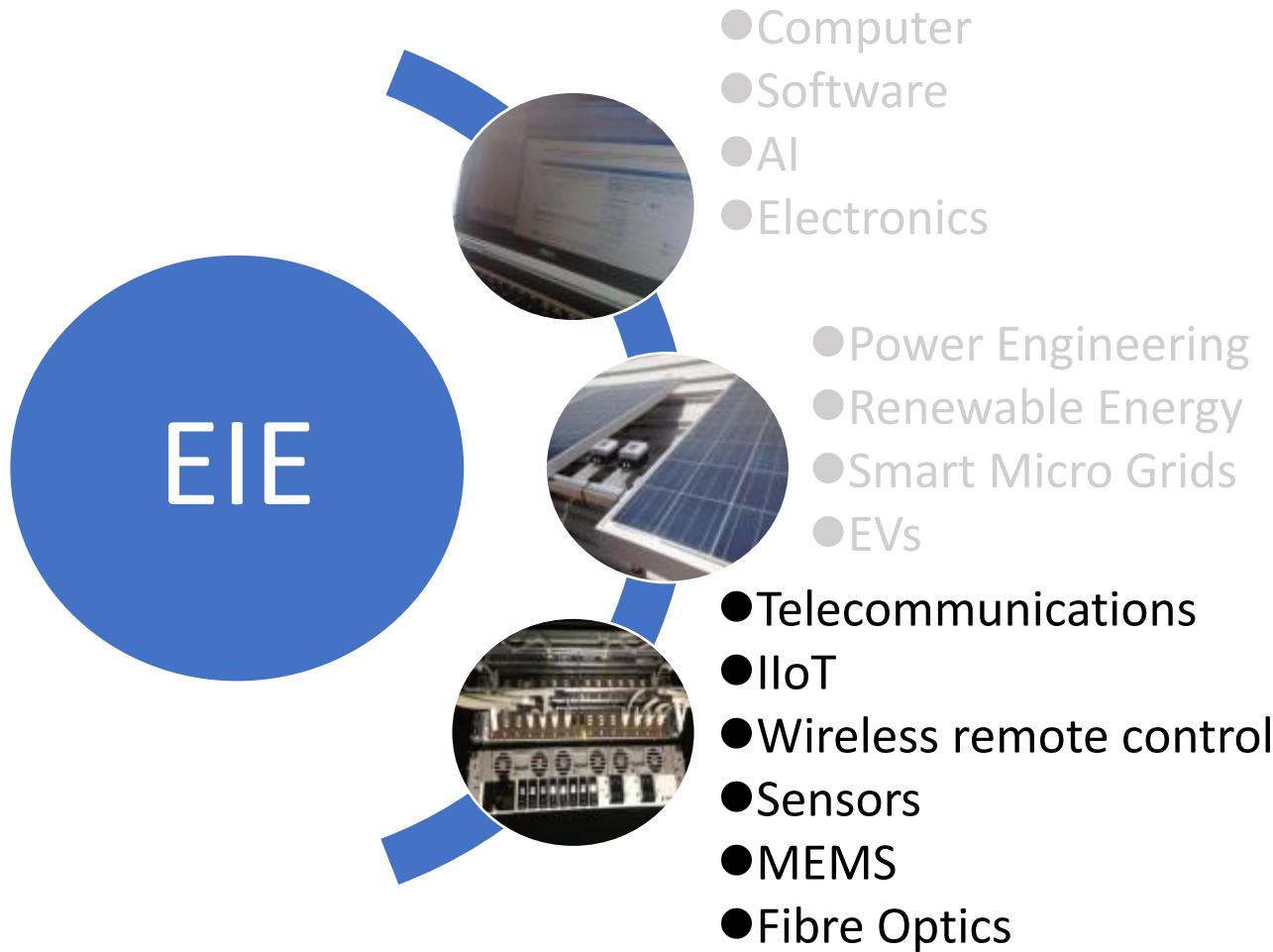


# What is Electrical and Information Engineering

- **Computer Engineering** (hardware, software engineering, artificial intelligence (AI), machine learning, big data, and cloud computing, etc.)
- **Intelligent Information and Software Engineering** (audio, video, VR, image processing, financial information)

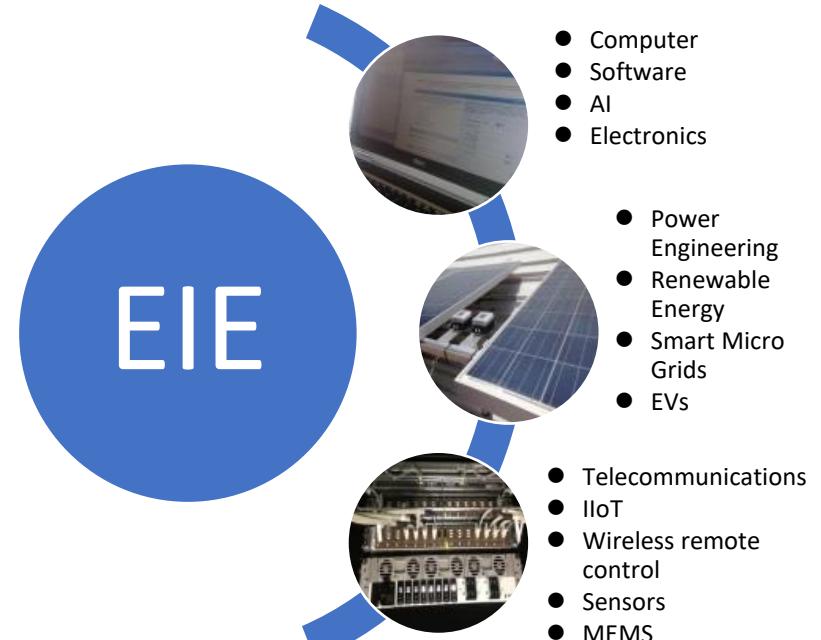


# What is Electrical and **Information** Engineering



# What is Electrical and Information Engineering

- **Telecommunications** (transmission of signals, images and sounds by wired/wireless, optical or other electromagnetic systems)
- **Internet of Things** (sensing, communications, data analytics & intelligence, and wireless remote control)



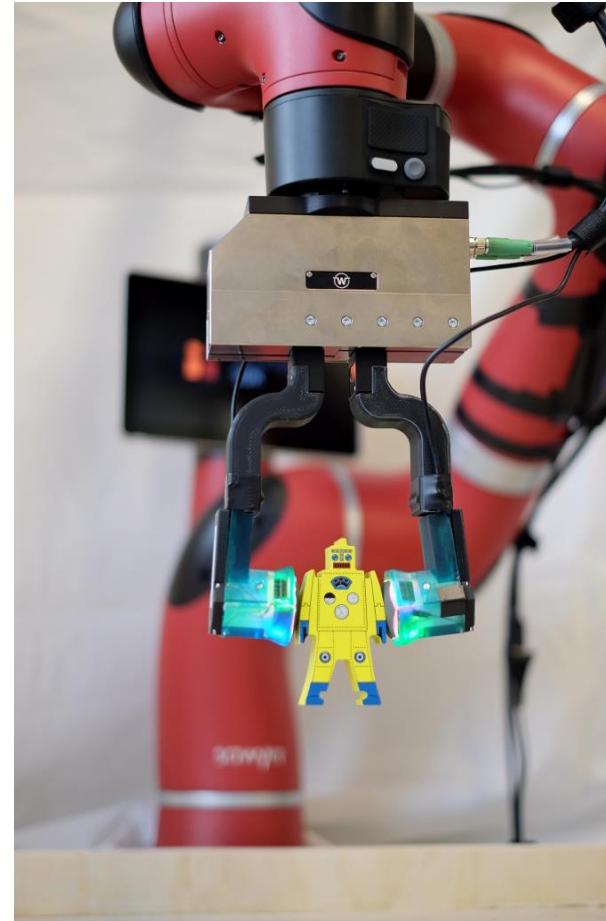
# 5G IIoT Testbed

- Implemented an M-CORD (Mobile Central Office Re-architected as a Datacentre) open-source programmable network platform running on general-purpose computers.
- The facility provide a programmable private network for a microgrid system, manufacturing and ecommerce warehouse automation with robot arms
- The SDN network platform and microgrid system have been built
- The manufacturing and warehouse automation are in progress



# IIoT Testbed

- The testbed consists of low cost robot arms and tactile sensors
- The robot arms are connected to the controller through wireless and wired networks
- Use cases: disassembly and cleaning of industrial equipment, handling products in ecommerce warehouses

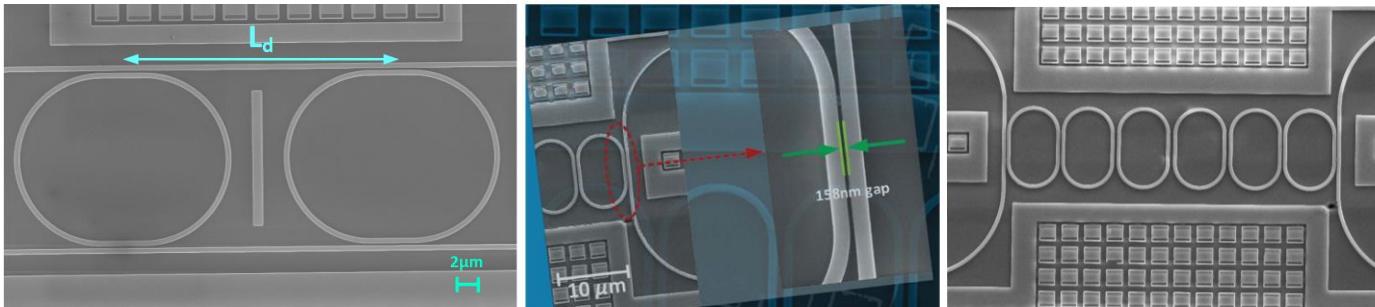
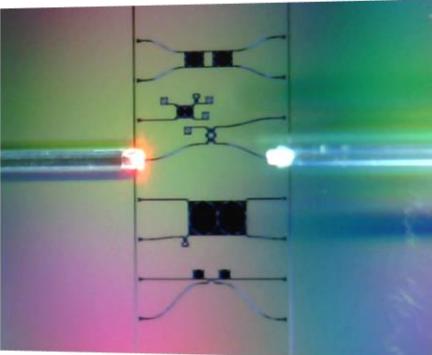
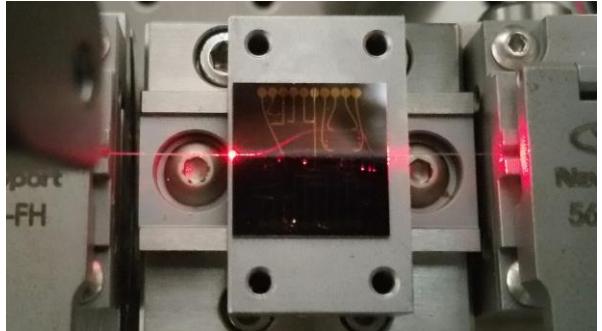


# SmartSat Project

- \$110 million 7 years' research project
- Low latency IoT architectures and technologies enabling universal IoT connectivity
- Integration of satellite and 5G systems
- IoT applications for remote safety, agriculture, mining, transport & logistics and environment monitoring



# Integrated Photonic Techniques for High Speed Signal Processing and Sensing

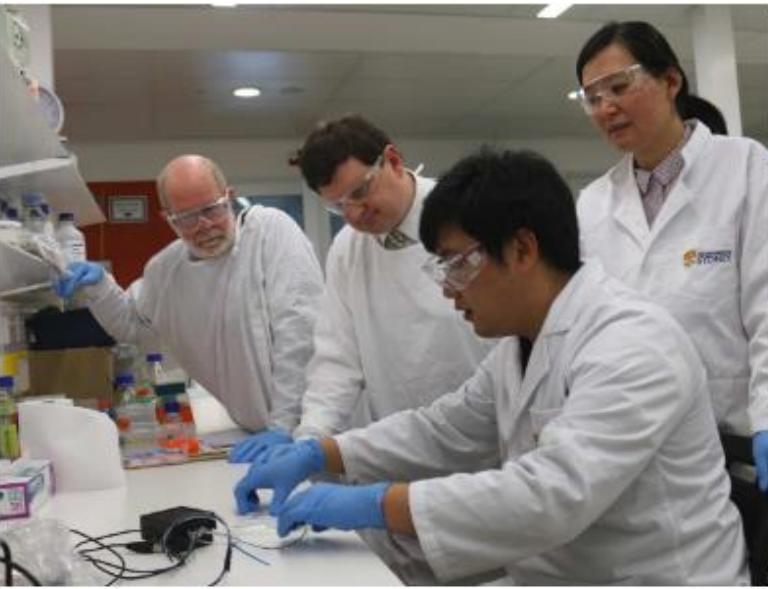
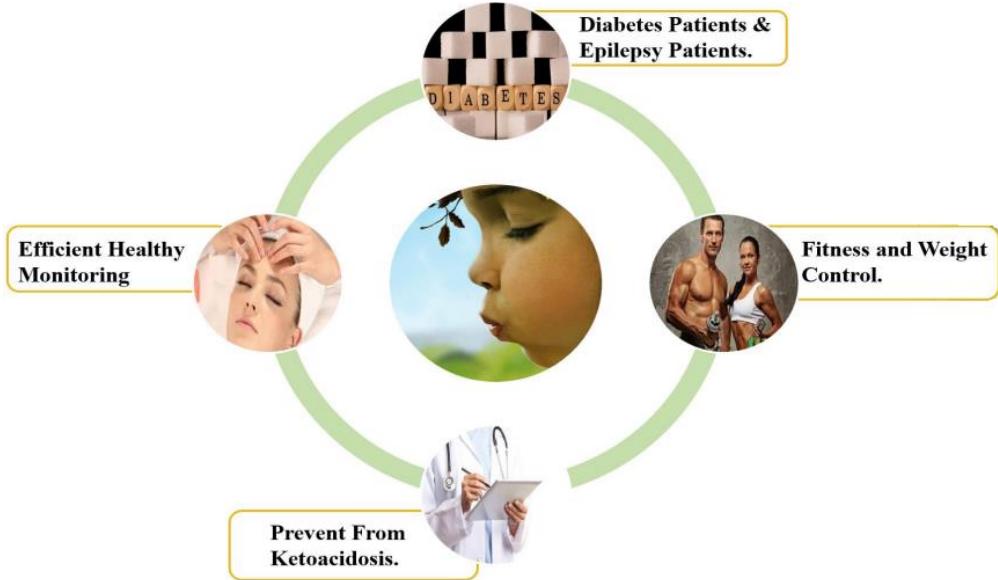


**Our Device Lab in Sydney Nano Institute**  
**COMMENT: CLEAN ROOM OF THE INSTITUTE**

- Integrated photonic circuits
  - CMOS-compatible fabrication
  - Cost-effective
  - Reduced footprint, power and weight

- Applications
  - Photonic assisted Radar
  - Lidar for autonomous vehicles
  - Sensors for various applications  
e.g. structural health monitoring and biomarker detection

# Biophotonics: Non-invasive Diabetes Detection



## Detecting diabetes' deadly ketones

A simple, hand-held breath testing device that detects deadly ketones has been developed by University of Sydney researchers.



Two lunch box size.



Handheld device.



Integration with smart phone and wearable devices.

(The technology is under patent licensing.)

- Truly non-invasive testing.
- Testing Ketone level from breath.
- Needle-free, risk-free and pain-free.
- Fast and high accuracy detection.
- Ultra-low cost
- Knowing your state of health any time anywhere.

### Key Researchers:

Prof Xiaoke Yi, School of Electrical and Information Engineering.

Prof Stephen Twigg, Charles Perkins Centre, Sydney Medical School, and Royal Prince Alfred Hospital.

A/Prof Paul Williams, University of Sydney and Royal Prince Alfred Hospital.

# Engineering Sydney Industry Placement Scholarship (ESIP Scholarship)

- Place our highest achieving students studying either an undergraduate engineering degree or Master of Professional Engineering within leading industry companies during their final year of studies.
- Your financial commitment
  - \$29,000 per student for each six-month placement.

Contact me for more details:  
[wanli.ouyang@sydney.edu.au](mailto:wanli.ouyang@sydney.edu.au)



# Welcome for collaboration

- Direct research funding
- Government funding, ARC Linkage, CRC, CRC-P
- Guest lecture
- ESIP Scholarship
- Final year projects

Contact: [wanli.ouyang@sydney.edu.au](mailto:wanli.ouyang@sydney.edu.au)

Thank you!  
I thank you!

Contact: wanli.ouyang@sydney.edu.au

Q&A

Contact: [wanli.ouyang@sydney.edu.au](mailto:wanli.ouyang@sydney.edu.au)