Karlsruher Institut für Technologie

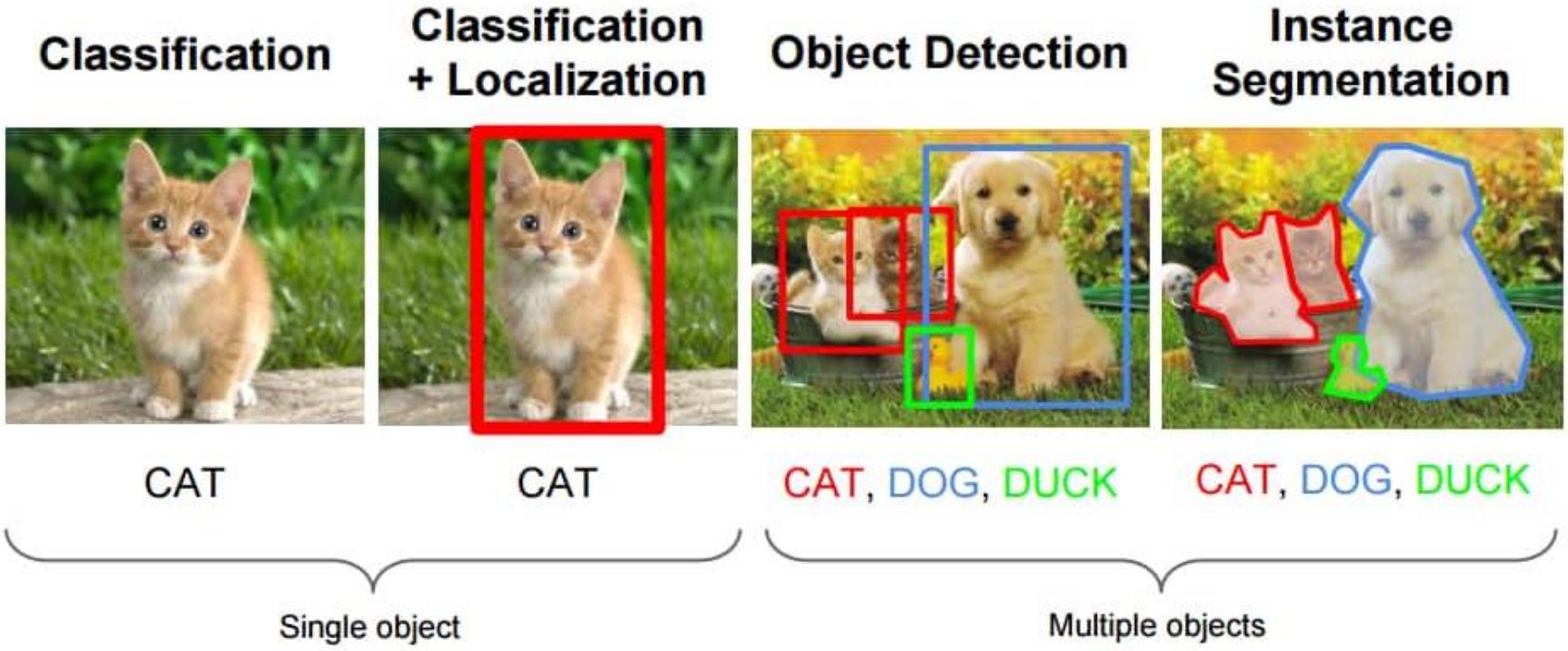
Deep Learning in der Bildverarbeitung

- Durchbrüche und Herausforderungen

Prof. Dr.-Ing. Rainer Stiefelhagen
Institute for Anthropomatics and Robotics (IAR) &
Study Centre for the Visually Impaired (SZS), KIT

The Deep Learning Revolution.

- In 2009 a Competition was made public in the Computer Vision field.
- It's name: ImageNet Large Scale Visual Recognition Competition.



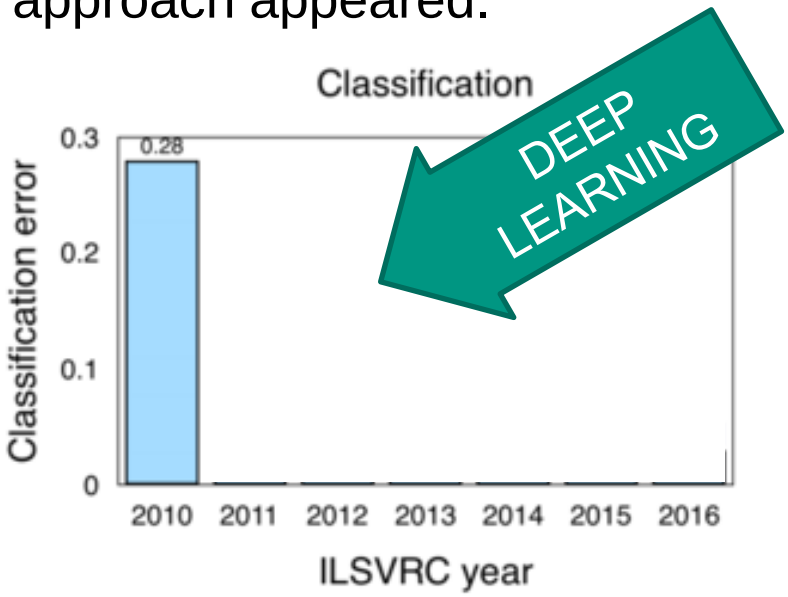
- And it was incredibly challenging.

ILSVRC data set

- 1000 classes, no class is an ancestor of another class
- Fine-grained categories (~120 types of dogs)

ImageNet Large Scale Visual Recognition Competition (ILSVRC)

- All submitted algorithms were extremely complicated, and its performance was not great. But then in 2012, a Deep Learning approach appeared.



ImageNet Classification with Deep Convolutional Neural Networks
Alex Krizhevsky, et Al.
NIPS 2012

- It was not only better than the competition but also simpler.

Deep Learning Increases Performance

Image Classification Winners:

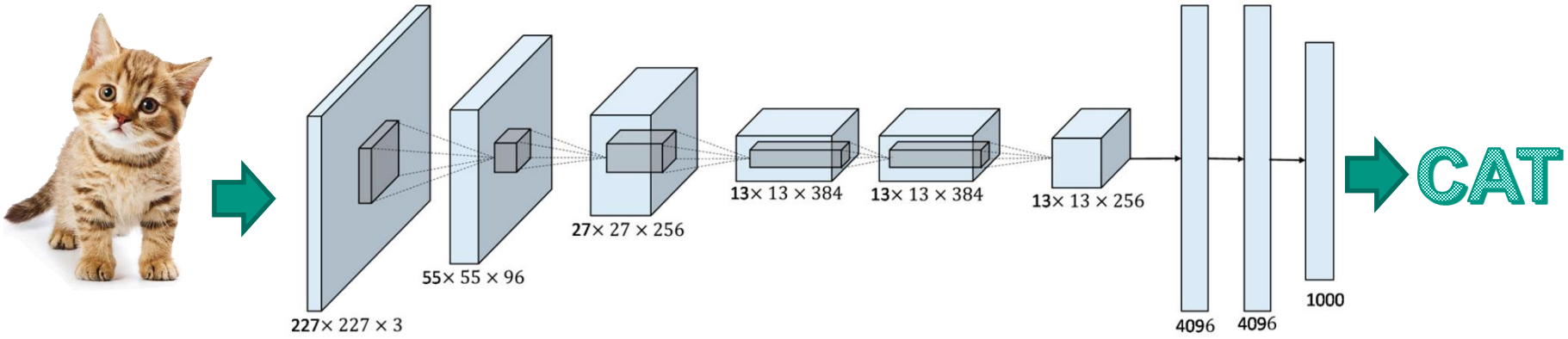
2012 Teams	%err
Alexnet (Toronto)	**15.3**
ISI (Tokyo)	**26.1**
VGG (Oxford)	26.9
XRCE/INRIA	27.0
UvA (Amsterdam)	29.6
INRIA/LEAR	33.4

2013 Teams	%err
Clarifai NYU spinoff	**11.7**
NUS (Singapore)	12.9
Zeiler-Fergus (NYU)	13.5
A. Howard	13.5
OverFeat (NYU)	14.1
UvA (Amsterdam)	14.2
Adobe	15.2
VGG (Oxford)	**15.2**
VGG (Oxford)	**23.0**

2014 Teams	%err
GoogLeNet	**6.6**
VGG (Oxford)	7.3
MSRA	8.0
A. Howard	8.1
DeeperVision	9.5
NUS-BST	9.7
TTIC-ECP	10.2
XYZ	11.2
UvA	**12.1**

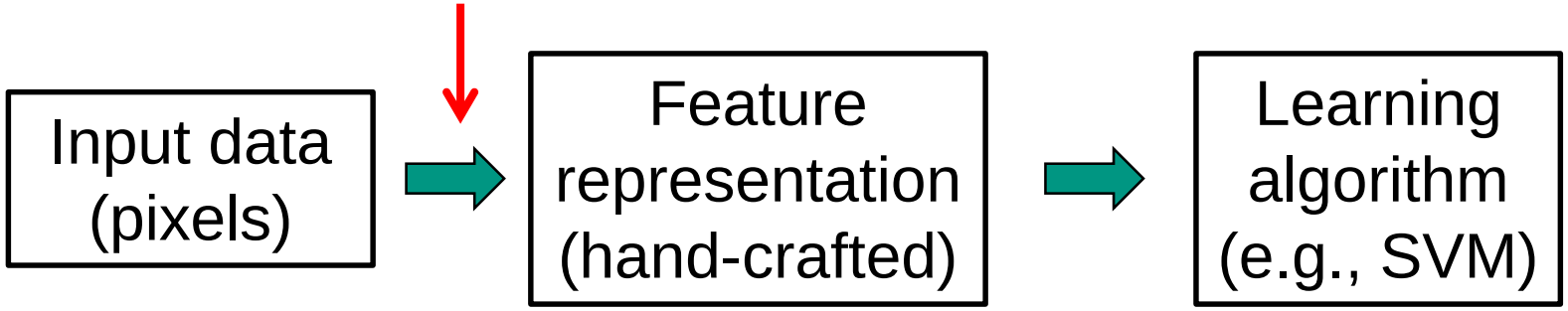
RED: Deep learning
Blue: Not deep learning (shallow)

Deep Learning is Simpler: End-to-end learning.



Traditional computer vision

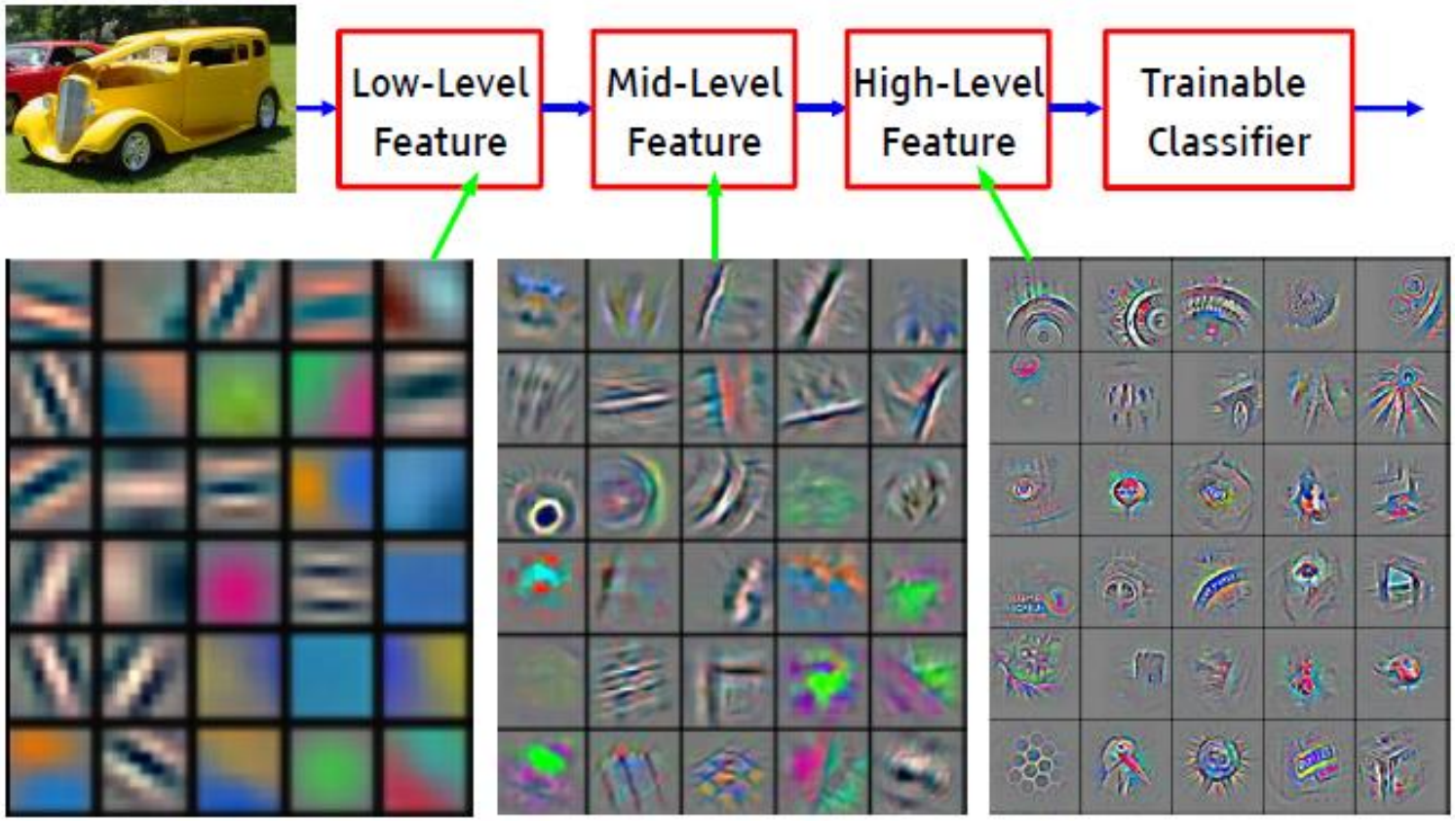
Features are not learned!



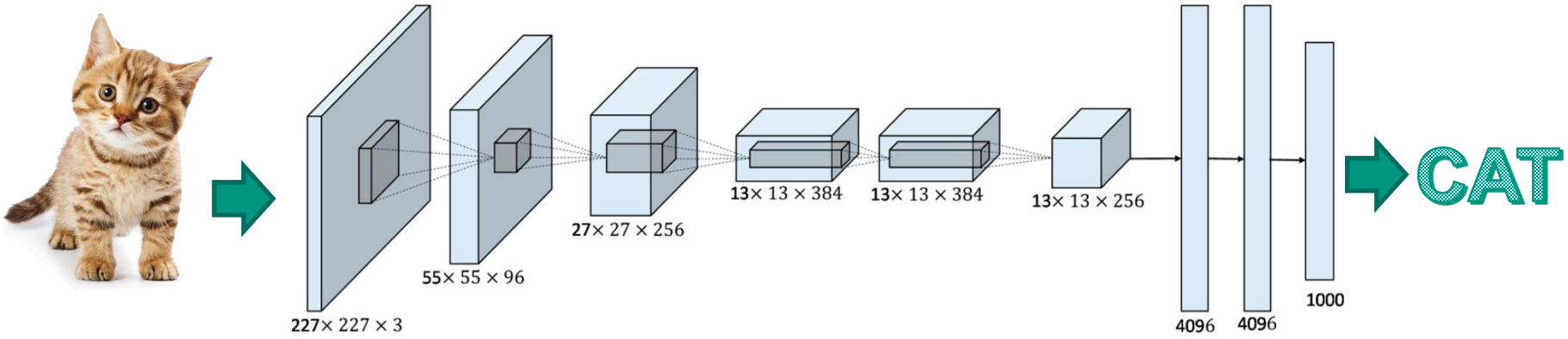
Stapler

Learning feature hierarchy (2)

- Learn a hierarchical representation of data



Deep Learning is Simpler: End-to-end learning.



Moderen CNN revolution

Large Scale Visual Recognition challenege (ILSVRC) winners

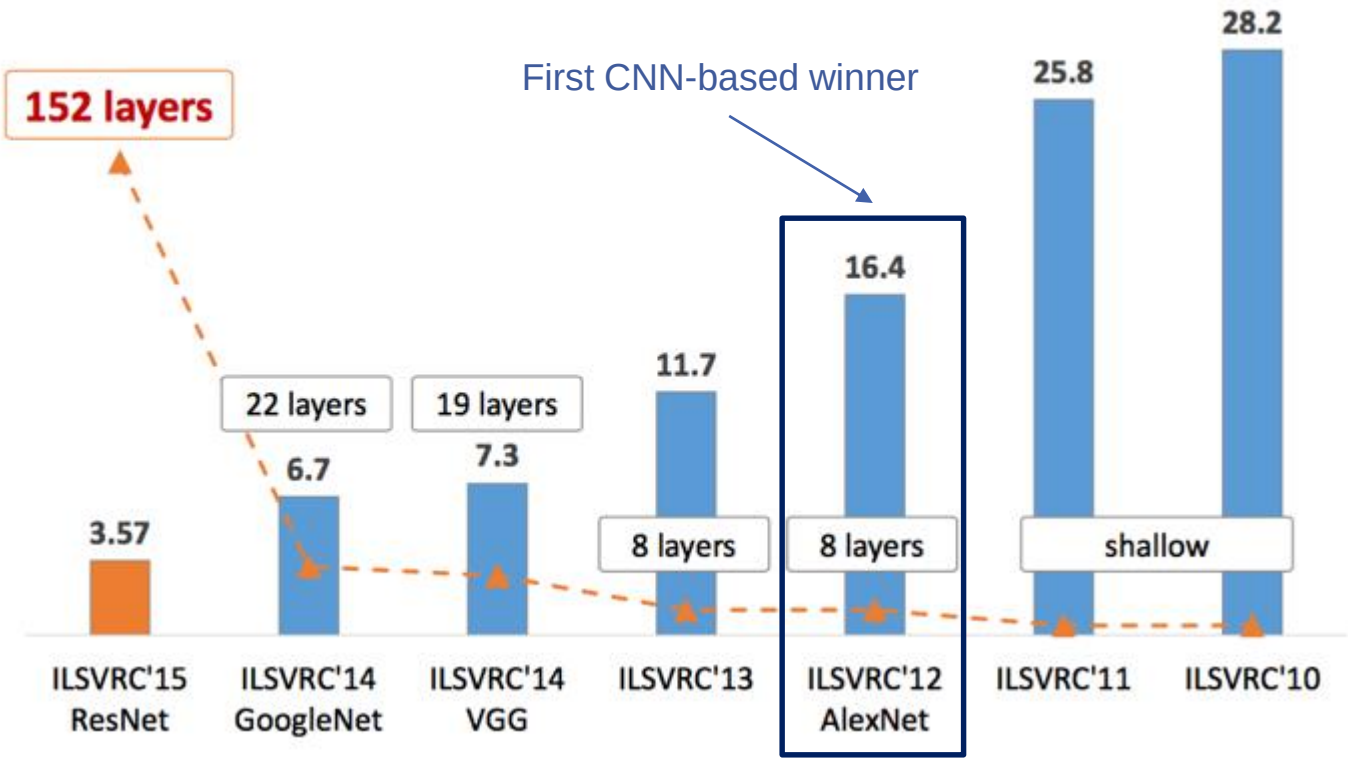


Figure copyright : Keiming He 2016

Simpler and Better, where is the catch?

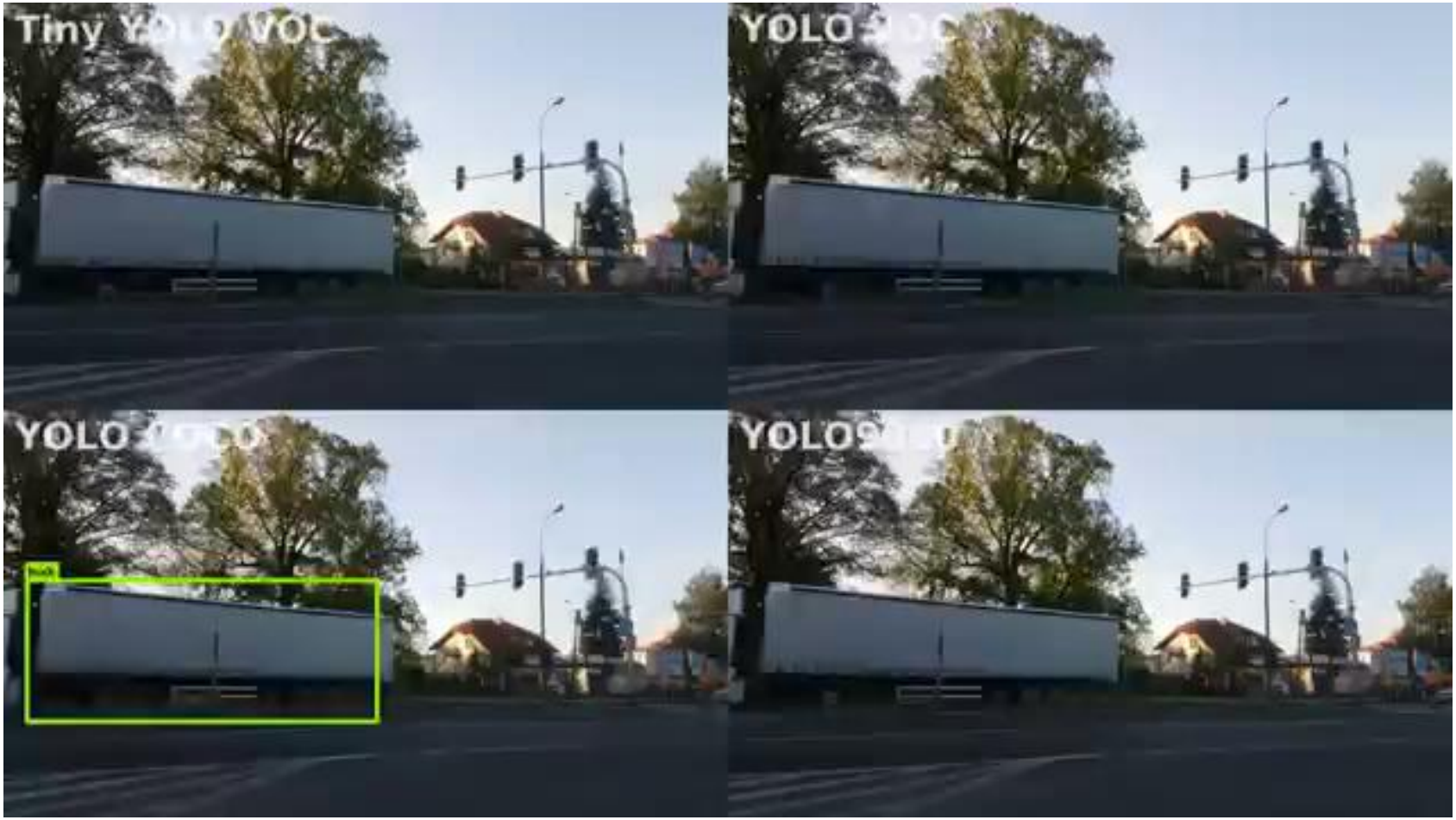
- The method is not new (i.e., LeCun, 1989)

- Deep Learning requires:
  - Large amounts of training data (i.e., ImageNet is 1000000 images).
  - Data must be annotated.
  - Training must be performed in GPUs, and takes days / weeks.

- Also, incremental progress does not work:
  - Few data, few annotations or few GPU time → No results.

Applications:

Fast Object Localization
You Only Look Once



[Redmon et al. CVPR 2016, 2017]

Semantic Segmentation
SegNet

[Badrinarayanan et al. PAMI 2017]

Convolutional Pose Machines

[Cao et al. CVPR 2017]

View Specific Pedestrian Attribute Inference (2)

[Sarfraz et al. BMVC 2017]

View Specific Pedestrian Attribute Inference (3)

[Sarfraz et al. BMVC 2017]

Image Captioning

A person riding a motorcycle on a dirt road.

Two dogs play in the grass.

A yellow school bus parked in a parking lot.

"black and white dog jumps over bar."

"young girl in pink shirt is swinging on swing."

Image & Video Question-Answering

What color are her eyes?
What is the mustache made of?

Does it appear to be rainy?
Does this person have 20/20 vision?

S. Antol, et al. VQA: Visual Question Answering. In *ICCV* 2015.

LOTR: Return of the King

Why does Arwen wish to stay in Middle Earth?

She is too weak to travel

She wants to die on Middle Earth

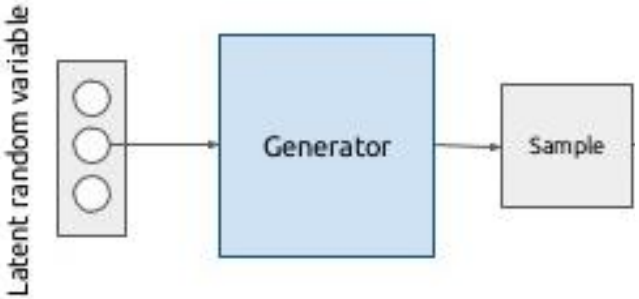
Her son asked her to stay

Arwen sees her son in her visions

She likes Middle Earth

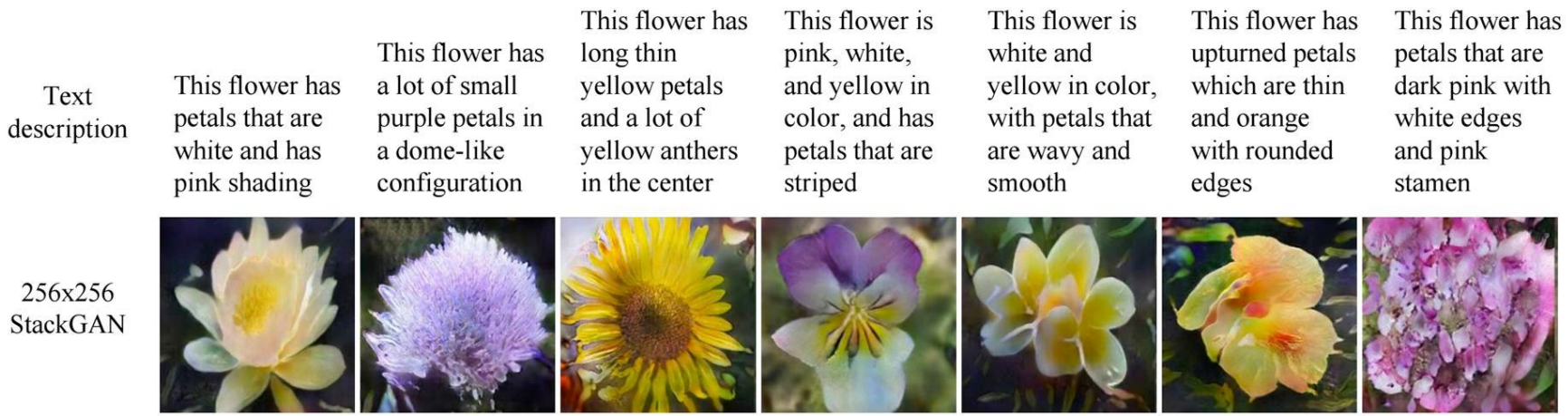
M. Tapaswi et al. MovieQA, In *CVPR* 2016.

Generative Adversarial Networks (GANs)



[Goodfellow et al. NIPS 2014]

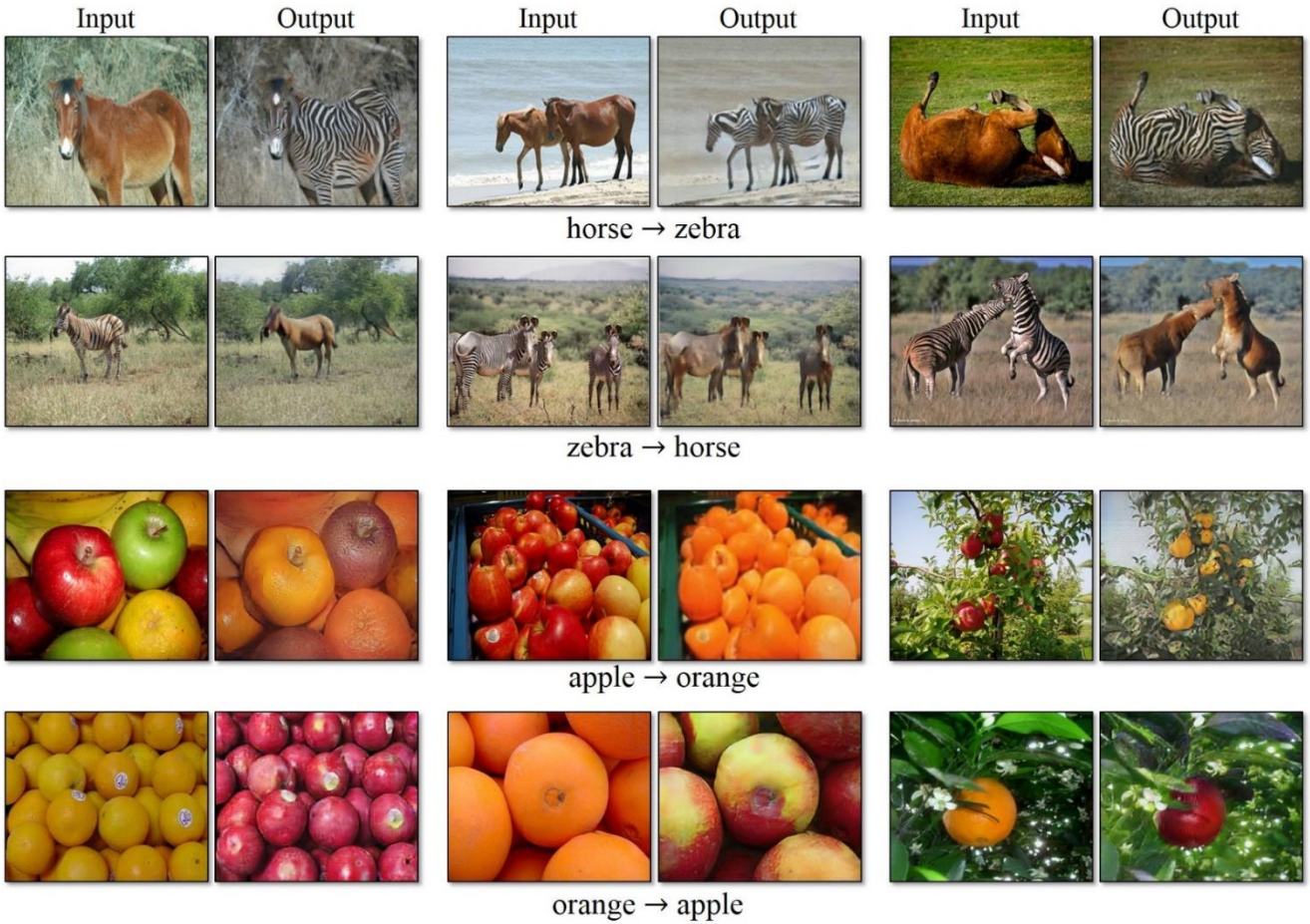
Generative Adversarial Networks (GANs)



[StackGAN, Zhang et al. ICCV 2016]

[DiscoGAN, Kim et al. ICML 2017]

Generative Adversarial Networks (GANs)



[CycleGAN, Zhu et al. ICCV 2017]

Challenges:

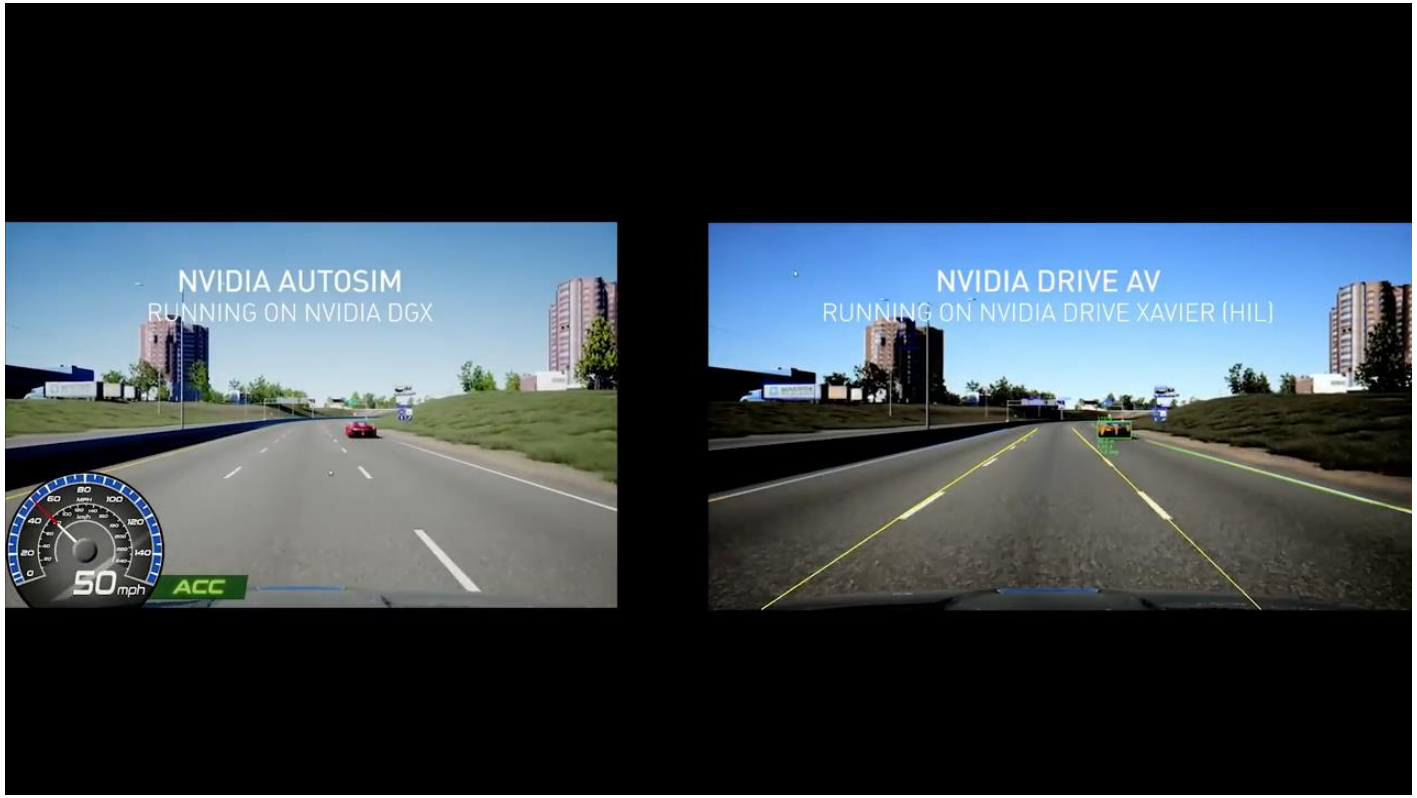
Dealing with the Unexpected:

- Deep Learning models learn statistics from the data. Unexpected or rare events, not present in the data, will not be recognized.

NVIDIA. CES 2016. e.g. Autonomous driving: how to avoid accidents if we have not enough data of them?

Dealing with the Unexpected:

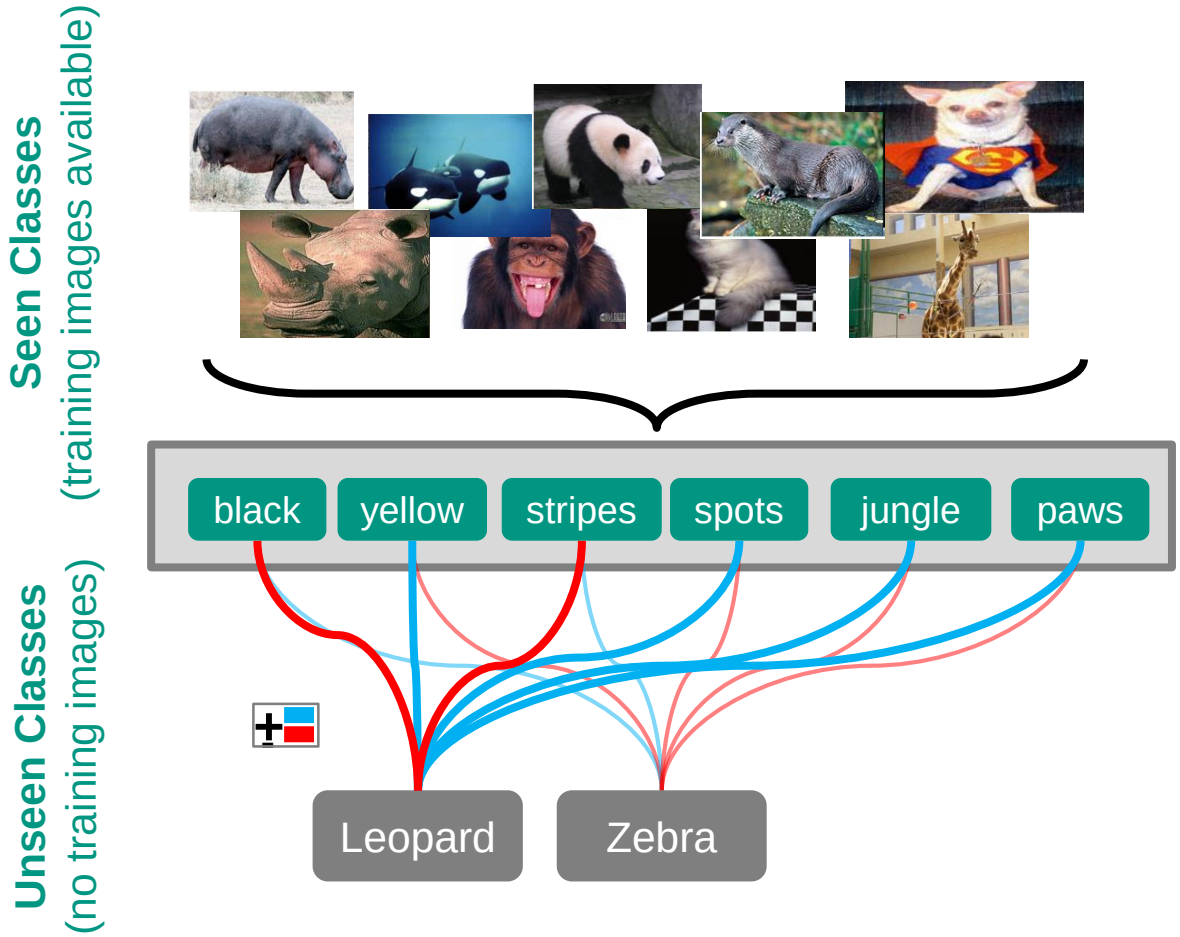
- Generate more data: (e.g., simulator, GANs)



NVIDIA. 2018. Driving Simulator.

Dealing with the Unexpected: Zero-Shot Learning

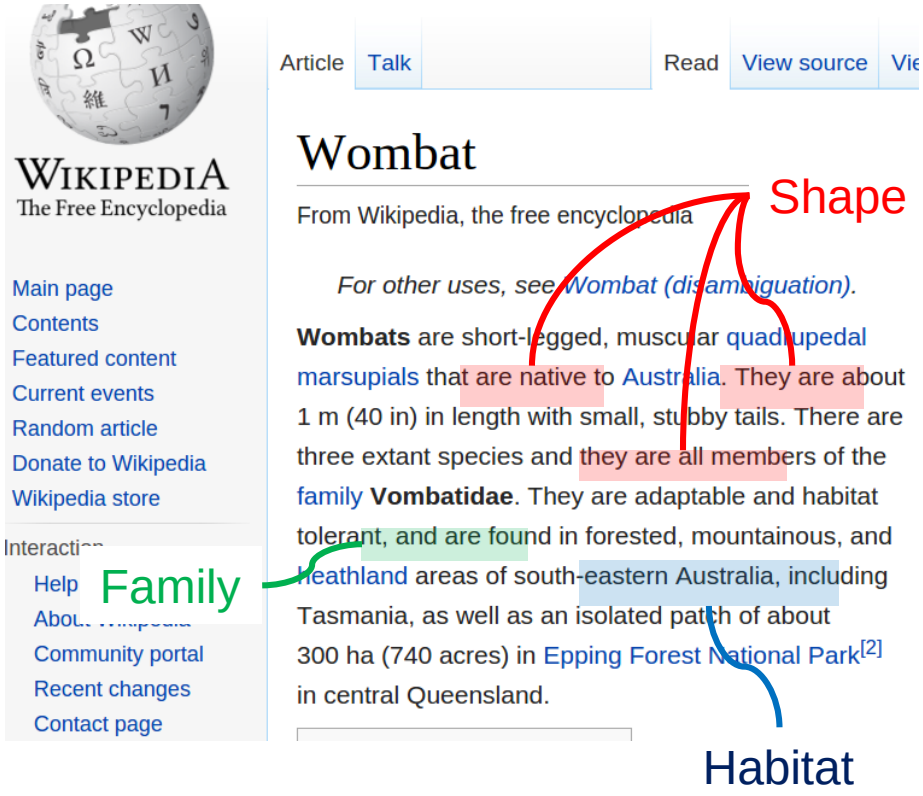
- Recognize objects with zero training data via knowledge transfer



[Z. Al-Halah et al., CVPR 2016]

Dealing with the Unexpected: Semantic Attribute discovery

- Online articles describing the classes
  - Automatic analysis → visual semantic attributes



[Z. Al-Halah et al., CVPR 2017]

More Challenges

- Deep Learning is a Black Box: how can it be trusted?
  - Bayesian Learning, Confidence Learning, etc.
- Deep Learning approaches are computationally expensive.
  - Custom Hardware is being developed (e.g., Intel Movidius Computing Stick)

- Lack of relevant amount of data.
  - Text/Image Knowledge bases.
  - Knowledge transfer.
- Not human-like intelligence: learns from examples, not from rules.
  - Rule-based machine learning.

Conclusions

- Deep Learning is a very powerful tool that can achieve close to human-like performance on some image analysis tasks.
- However, to use Deep Learning, an investment on GPUs, annotated data, and time is required.
- Also, being a statistical method (i.e., that learns from data), does not cope well with rare or unexpected events.
- New developments in transfer learning, GANs, and simulators are expanding the number of applications where Deep Learning can be applied.