ISTAR IUL
INFORMATION SCIENCES, TECHNOLOGIES AND ARCHITECTURE RESEARCH CENTER
UNIVERSITY INSTITUTE OF LISBON

Efficiency and Scalability of Multi-Lane Capsule Networks

Mauricio Breternitz[!] – ISCTE-IUL, ISTAR-IUL
mbjrz@iscte-iul.pt
Vanderson Rosario[!], UNICAMP
Edson Borin, UNICAMP

ISCTE IUL
Instituto Universitário de Lisboa

! in collaboration with INESC-ID / IST

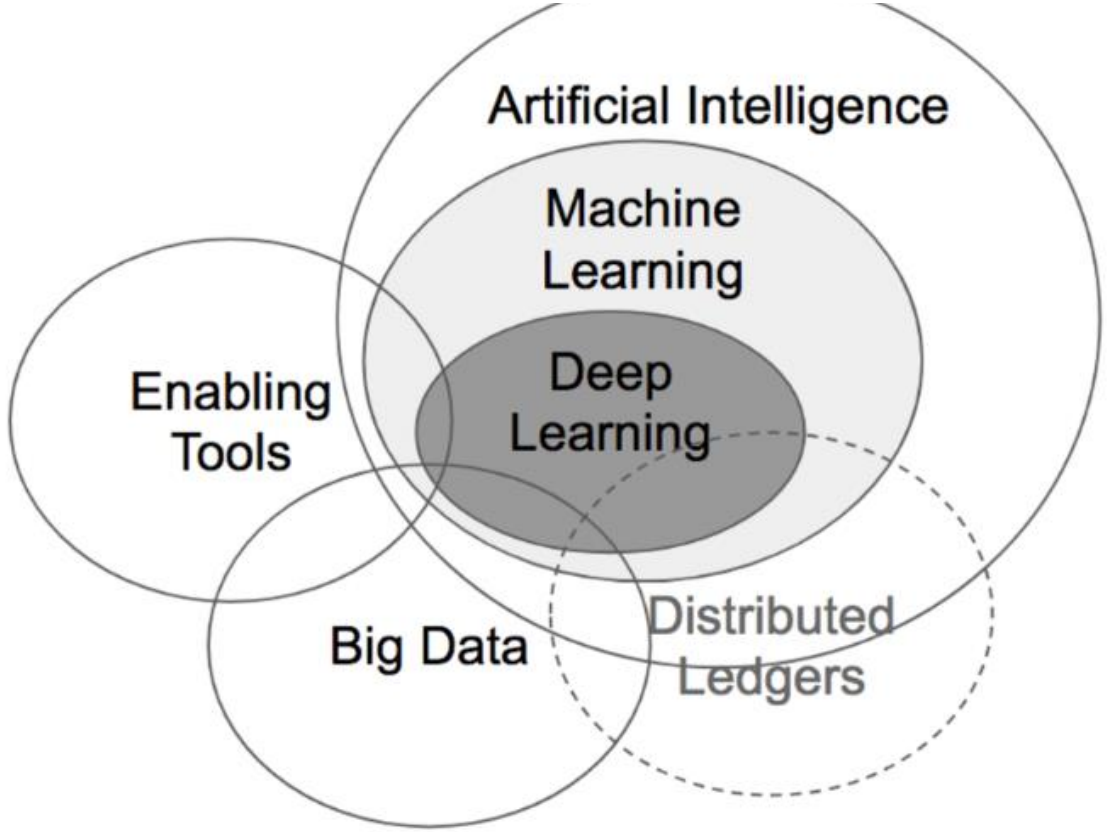
Outline

- Motivation – Smart Cities
- Deep Neural Networks
- Machine Learning
  - MLCN – Multi-Lane Capsule Network
- MLCN Operation and Results
- Next Steps

Forbes Jun 26, 2018, 01:22pm

ISTAR IUL
INFORMATION SCIENCES, TECHNOLOGIES AND ARCHITECTURE RESEARCH CENTER
UNIVERSITY INSTITUTE OF LISBON

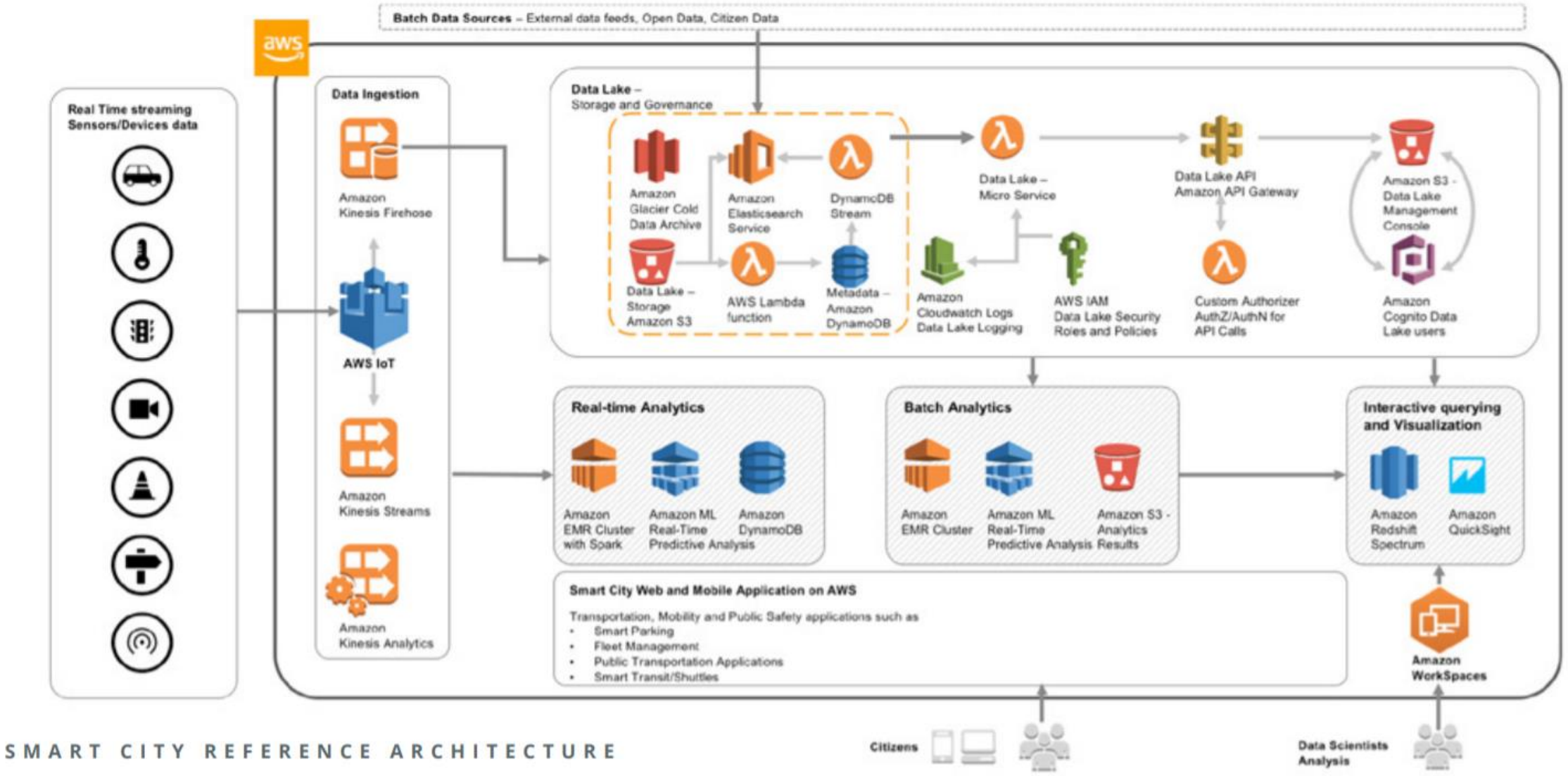
How Cities Are Getting Smart Using Artificial Intelligence



How AI is helping cities get smarter TOM VANDER ARK

https://www.forbes.com/sites/tomvanderark/2018/06/26/how-cities-are-getting-smart-using-artificial-intelligence/#239a0e1438

SMART CITY ARCHITECTURE

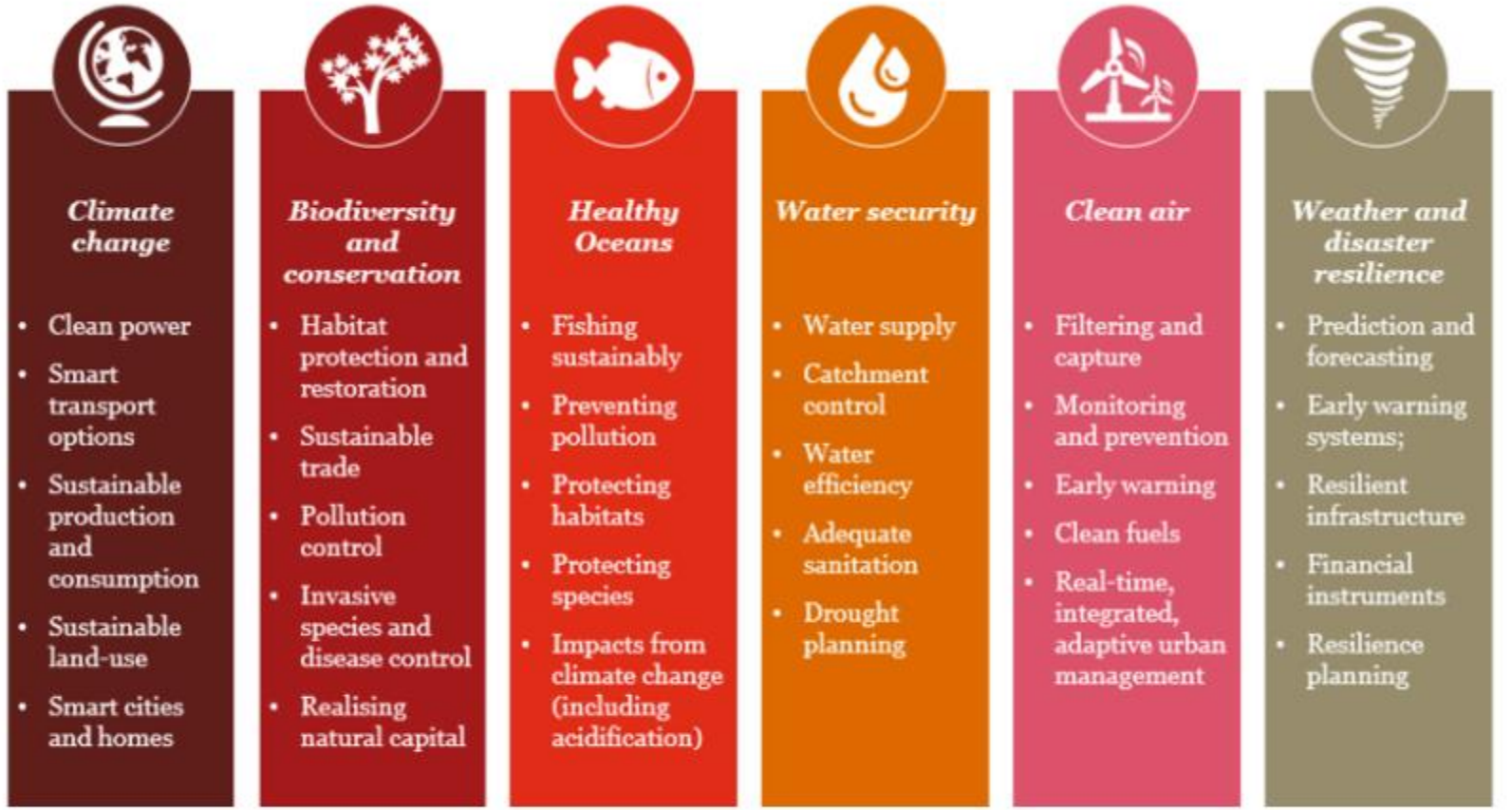


Source: Deloitte Digital, Smart Cities and the Journey to the Cloud, November, 2019

8 ways AI can help save the planet

ISTAR IUL
INFORMATION SCIENCES, TECHNOLOGIES AND ARCHITECTURE RESEARCH CENTER
UNIVERSITY INSTITUTE OF LISBON

WORLD ECONOMIC FORUM



Priority action areas for addressing Earth challenge areas

Image: PwC

https://www.weforum.org/agenda/2018/01/8-ways-ai-can-help-save-the-planet/

DEFINITION OF MACHINE LEARNING

- Simple Definition: "Algorithms that Learn From Data"

Traditional Programming

In Traditional Programming, a Human expert encodes his knowledge of the relationship of data and desired output as a program to process input data to generate the desired output

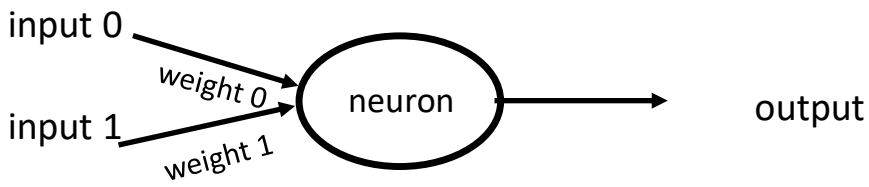
Machine Learning

In Machine Learning, the system autonomously learns the relationship of data and the desired output, creating classification rules (inference) to provide the desired output from similar input

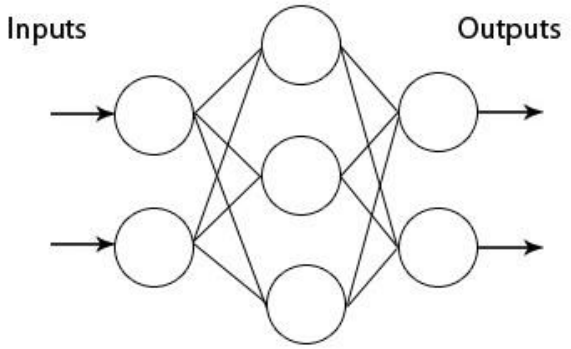
- Machine Learning: A system capable of the autonomous acquisition and integration of knowledge

DEEP NEURAL NETWORKS

- *rapidly becoming the preferred algorithm, currently the best solutions for image/speech/natural language processing*
- *Biologically-inspired: simulated neurons*
- *Good match for GPU acceleration because the mathematical operation to compute the effects of weighted inputs for multiple neurons is a matrix-vector multiplication.*

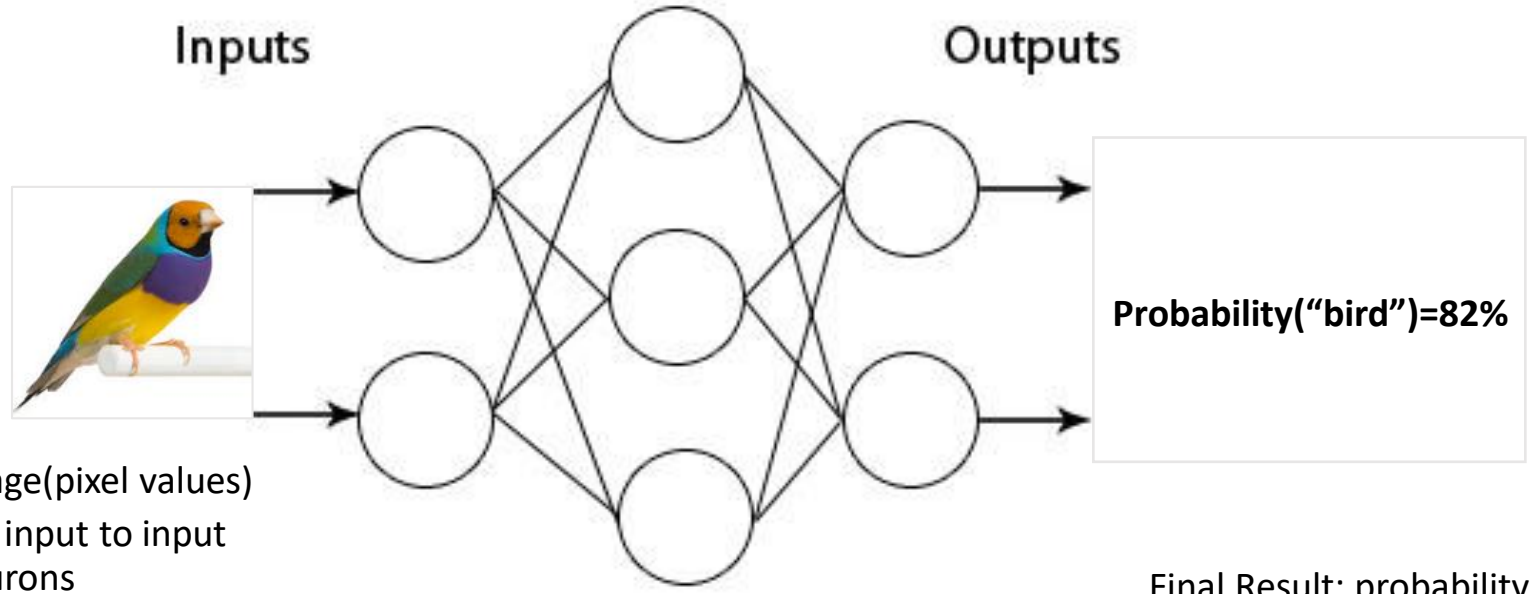


*A Simulated Neuron: A biologically inspired algorithm whereby a number of input values are provided to a simulated neuron, which computes an output based on a **weighted** combination of the input values*



An Example Deep Neural Network(DNN): A multi-layered sequence of simulated neurons

EXAMPLE:
DEEP NEURAL NETWORK CLASSIFYING AN IMAGE



Image(pixel values) are input to input neurons

Groups of neurons are trained to detect specific features in image regions.

Subsequent layers reacts to combinations of features that compose the image.

Final Result: probability of input image being a given entity

ML Challenges

-CNN challenges

-chihuahua, adversarial input

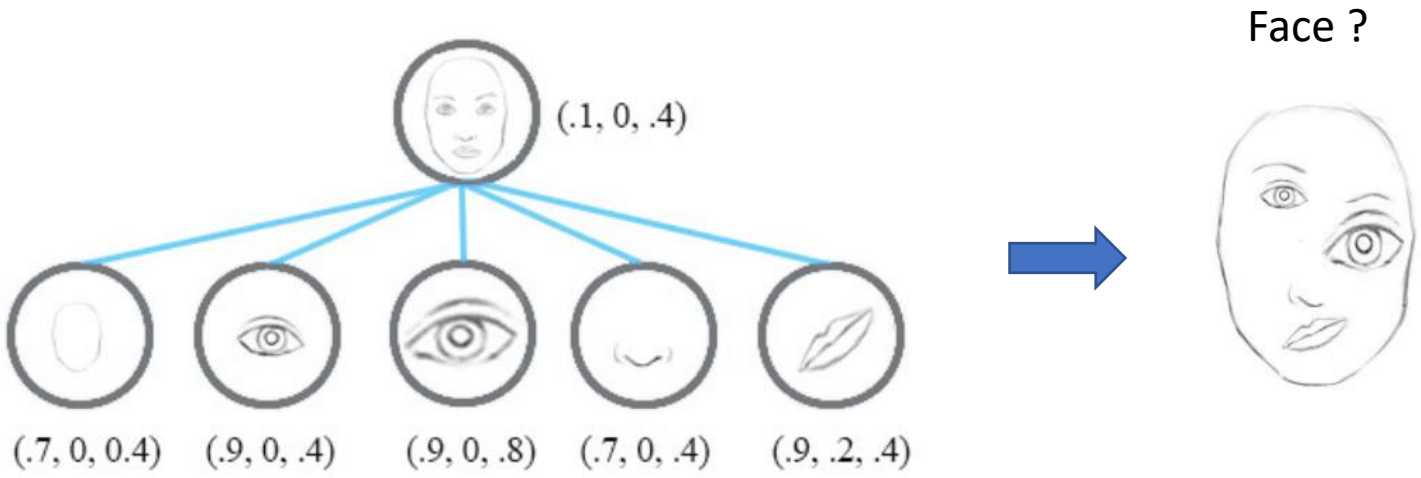
-Network Inspection

-Who does what?

Capsule Networks

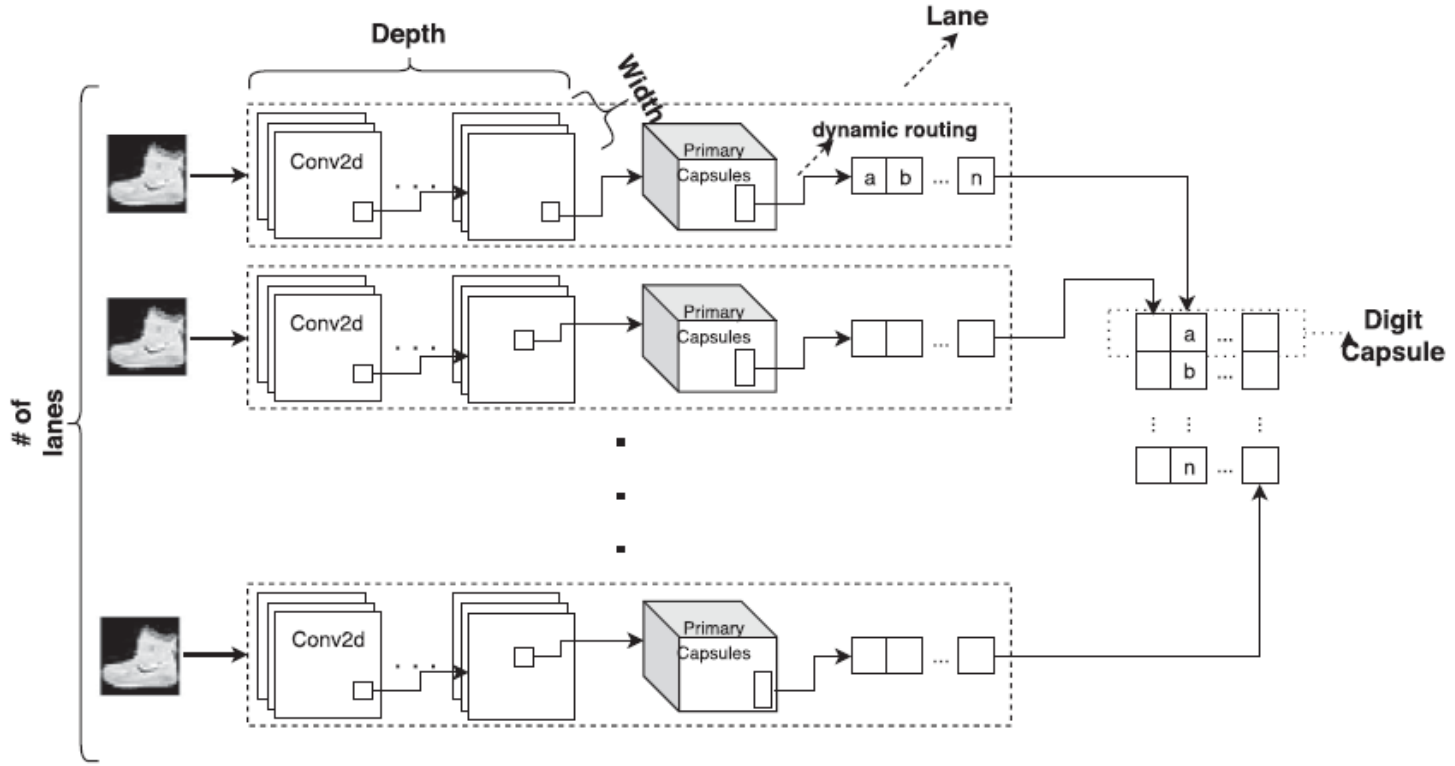
A capsule is a group of neurons that not only capture the ***likelihood*** *but also the* ***parameters*** *of the specific feature.*

Sabour, Sara, Nicholas Frosst, and Geoffrey E. Hinton. "Dynamic routing between capsules." *Advances in Neural Information Processing Systems*. 2017.



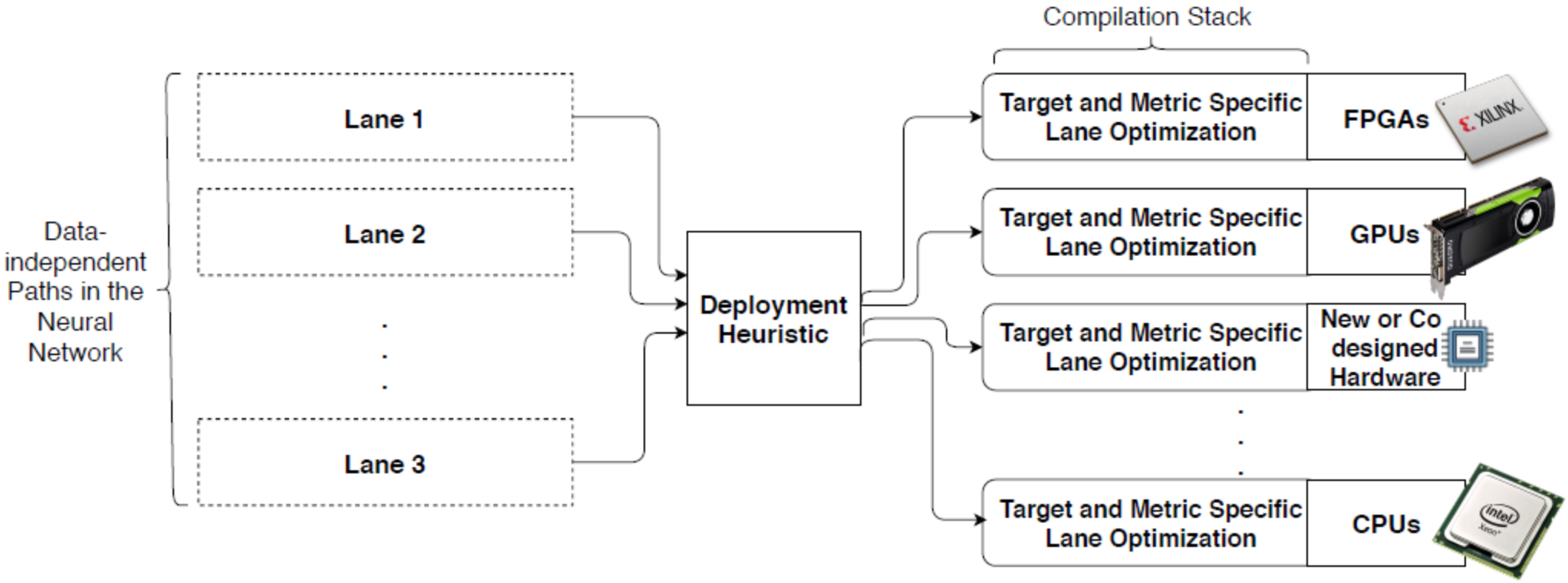
The Multi-Lane Capsule Network (MLCN)

ISTAR IUL
INFORMATION SCIENCES, TECHNOLOGIES AND ARCHITECTURE RESEARCH CENTER
UNIVERSITY INSTITUTE OF LISBON



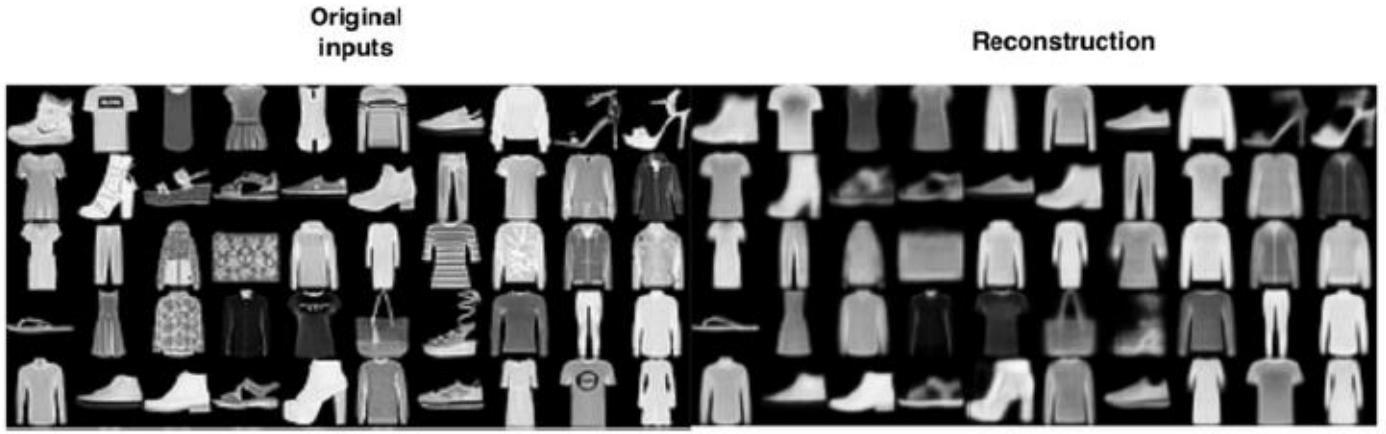
"The multi-lane capsule network,"
V. M. do Rosario ; Edson Borin ; Mauricio Breternitz,
IEEE Signal processing letters, vol. 26, pp. 1006–1010, 2019

Hardware Substrates For MLCN

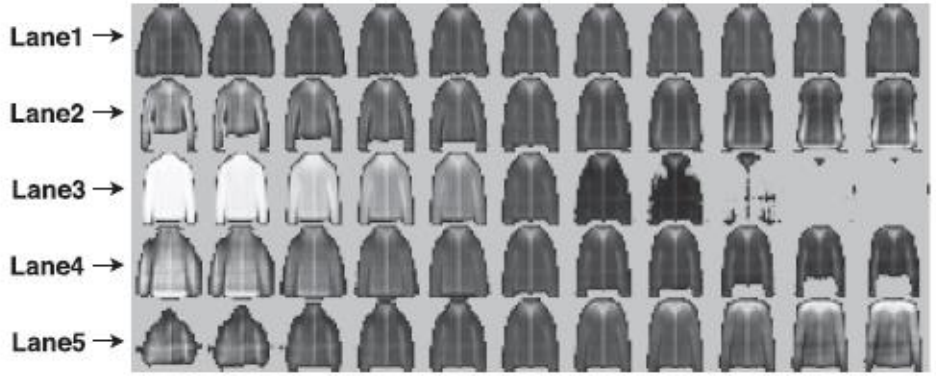


Multiple Neural Network *lanes* can be trained in parallel using multiple HW even in heterogeneous scenarios..

MLCN Operation



Reconstructions from the Fashion-MNIST using MLCN.



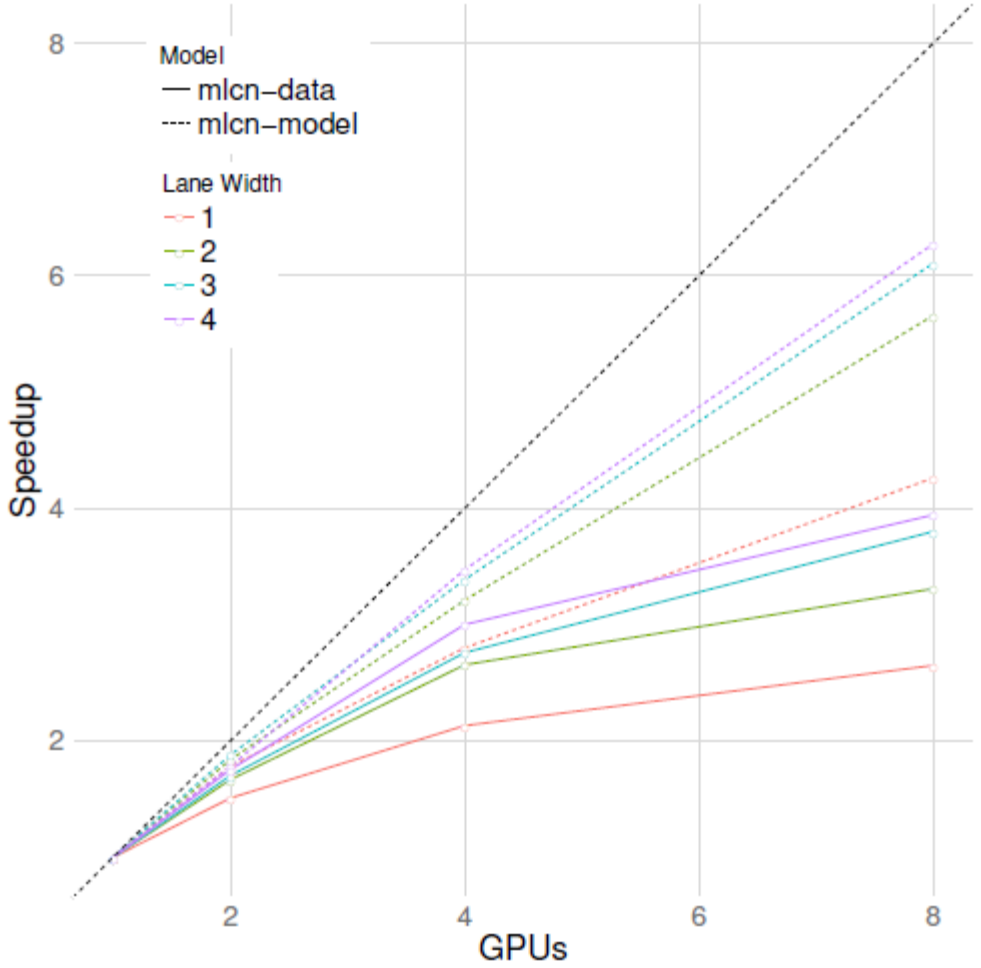
Synthetic variation on the lanes output.

MLCN Accuracy

TABLE I: Comparison between Baseline CapsNet and MLCN.

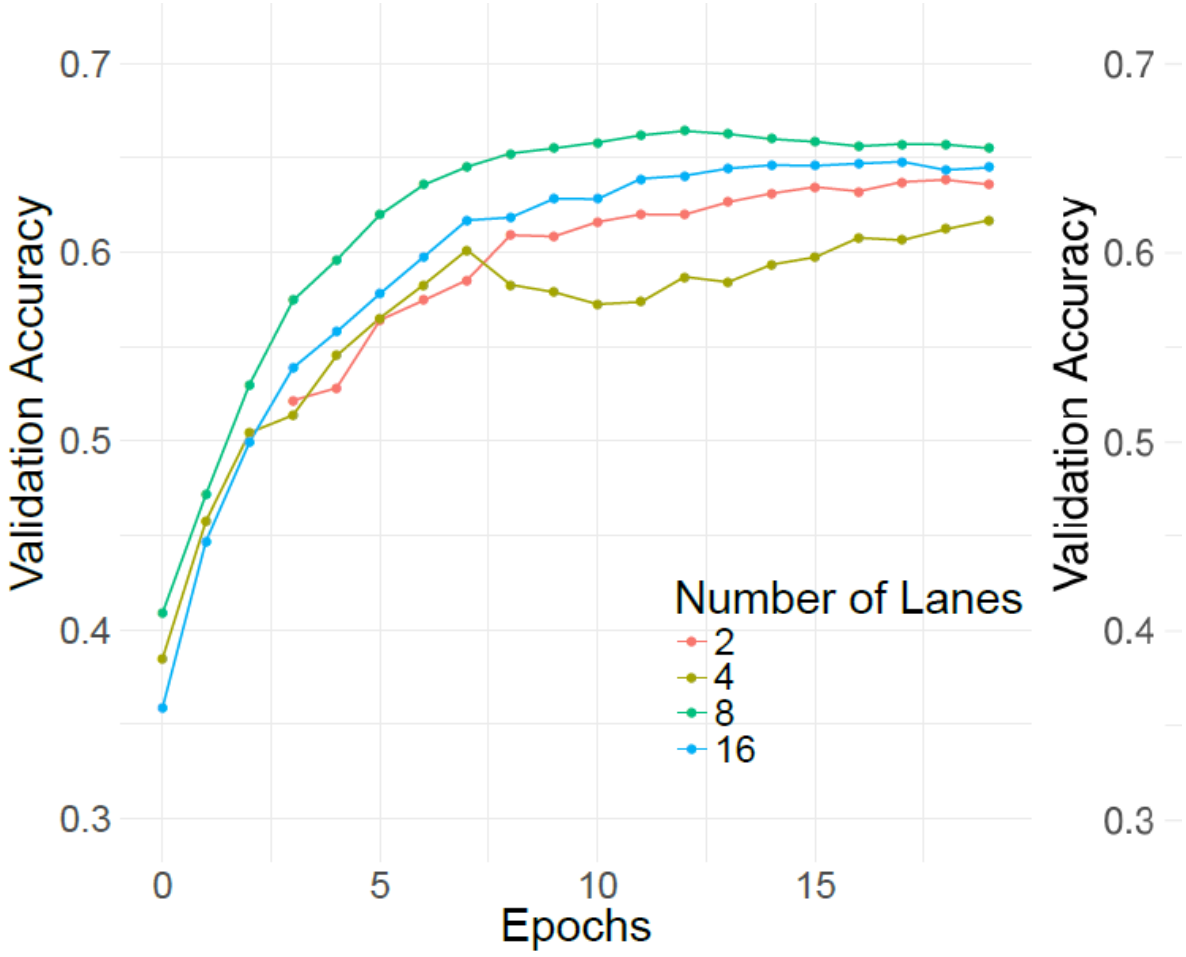
Network/set	# of *lanes*	*lane*'s Width	Params.	Train Time (sec./epoch)	Accuracy
Cifar10:					
Baseline	-	-	11k	240	66.36%
Mlcn2	4	4	5k	53	69.05%
Mlcn2	32	2	14k	204	75.18%
Fashion-MNIST:					
Baseline	-	-	8k	220	91.30%
Mlcn2	2	4	3.6k	20	91.01%
Mlcn2	8	4	10.6k	92	92.63%

MLCN speedup



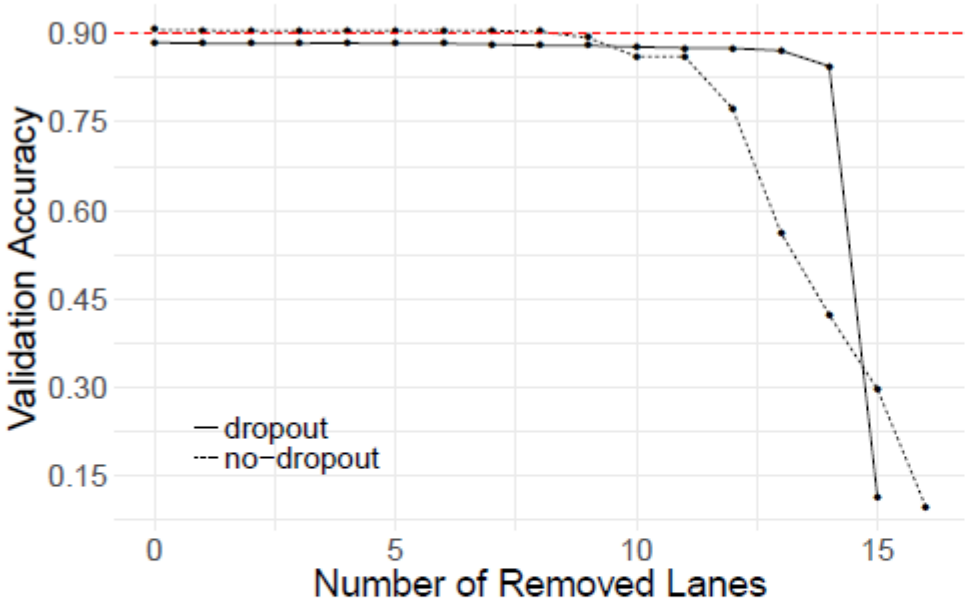
MLCN using model-parallelism with mini batch width of 150 and varying the width of the *lanes*.

Varying the Number of Lanes

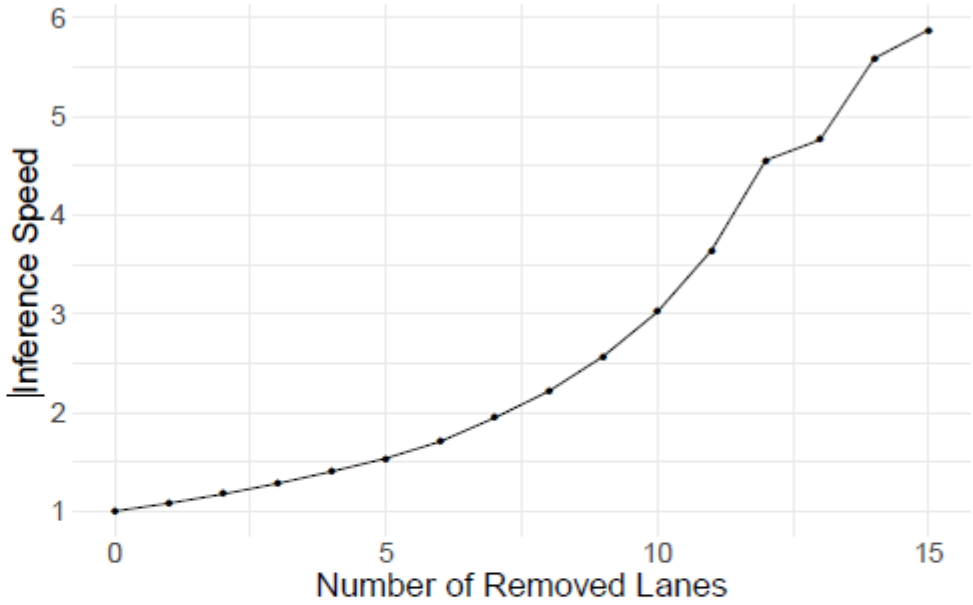


(c) Mlcn1 - Cifar10

Useful Lanes



(a) Accuracy



(b) Speed

Impact of removing lanes in order, starting with the less useful to the most.

Next Steps

- Efficient MLCN deployment
  - Compiler-based framework
- CapsNet-based IoT
  - Smart cities/sustainability

- Contact: mbjrz@iscte-iul.pt

References

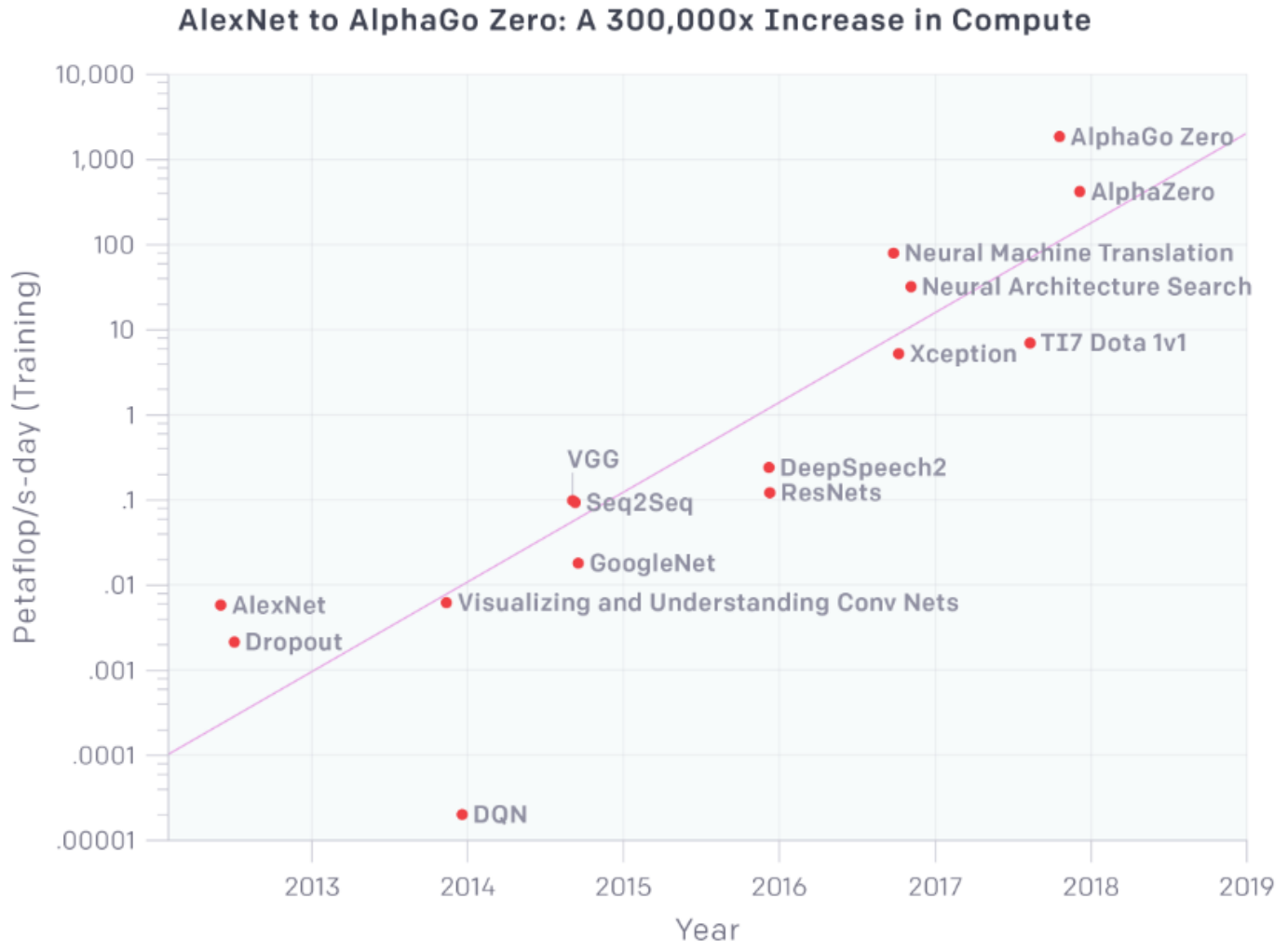
- do Rosario, V. M., Borin, E., & Breternitz, M. (2019). The Multi-Lane Capsule Network. *IEEE Signal Processing Letters*, *26*(7), 1006-1010.

- do Rosario, V. M., Borin, E., & Breternitz, M. (2019). The Multi-Lane Capsule Network. *IEEE Signal Processing Letters*, *26*(7), 1006-1010.

do Rosario, V. M., Borin, E., & Breternitz, M. (2019).
The Multi-Lane Capsule Network. *IEEE Signal Processing Letters*, *26*(7), 1006-1010.

AI Increasing Demand for Computational Power

ISTAR IUL
INFORMATION SCIENCES, TECHNOLOGIES AND ARCHITECTURE RESEARCH CENTER
SITY INSTITUTE OF LISBON



According to OpenAI, the demand for compute by deep learning networks has been doubling every 3.5 months since 2012.

Source: https://www.zdnet.com/article/ai-is-changing-the-entire-nature-of-compute/

ISTAR IUL
INFORMATION SCIENCES, TECHNOLOGIES
AND ARCHITECTURE RESEARCH CENTER
UNIVERSITY INSTITUTE OF LISBON

ISCTE IUL
Instituto Universitário de Lisboa