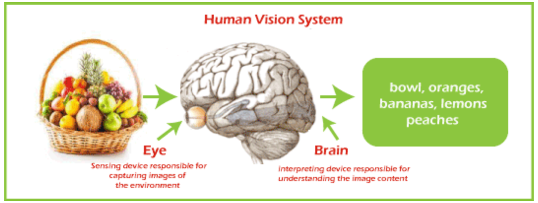
MPSoC'25

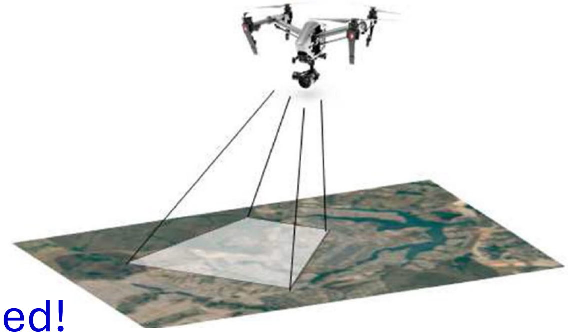
23rd MPSoC

Les Fermes de Marie, Megeve, France, June 15th-20th, 2025

KIOS Research and Innovation Center of Excellence

Dynamic Convolutional Neural Networks for Embedded Computer Vision





Do only ***what*** and ***when*** you need!

Theocharis (Theo) Theocharides

Associate Professor & Department Head, Department of Electrical and Computer Engineering

Director of Research, KIOS Research and Innovation Center of Excellence

University of Cyprus

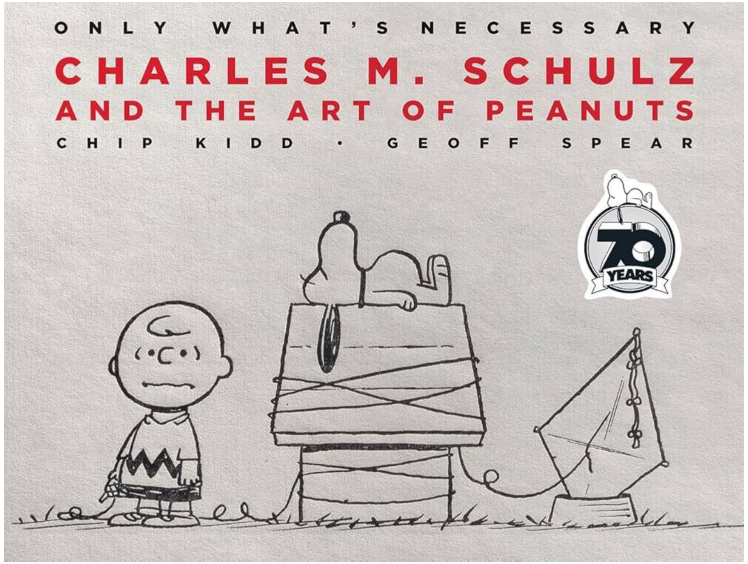
University of Cyprus

Imperial College London

This work is supported by the European Union's Horizon Europe research and innovation programme under grant agreement No 101168067 (GuardAI) and the from the European Union's Horizon 2020 research and innovation programme under grant agreement No 739551 (KIOS CoE) & from the Government of the Republic of Cyprus through the Cyprus Deputy Ministry of Research, Innovation and Digital Policy.

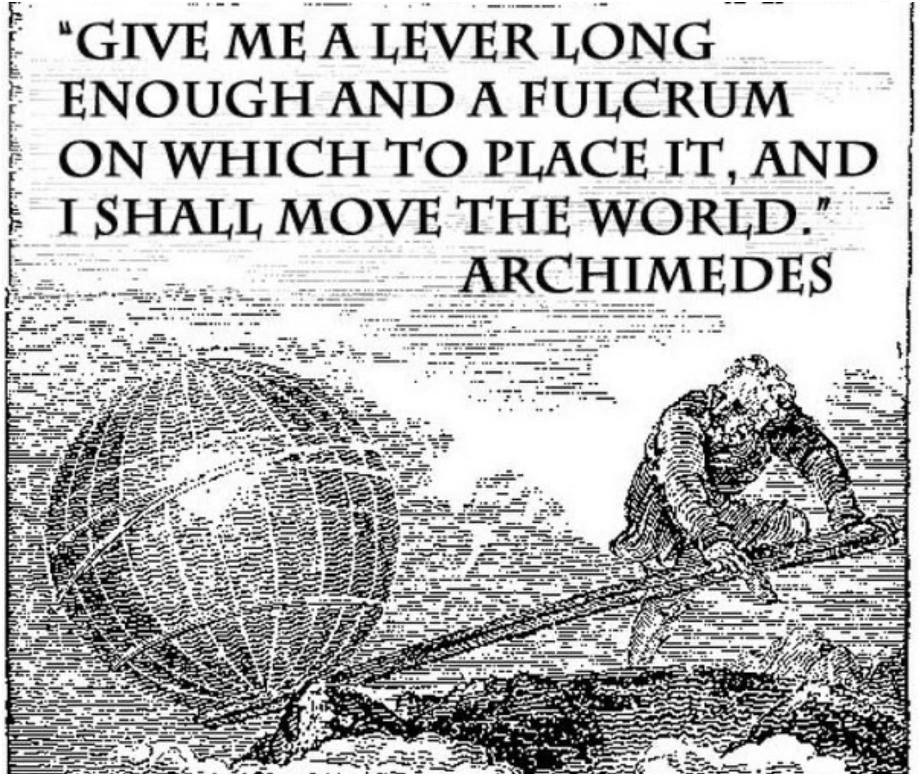
FUNDED BY:

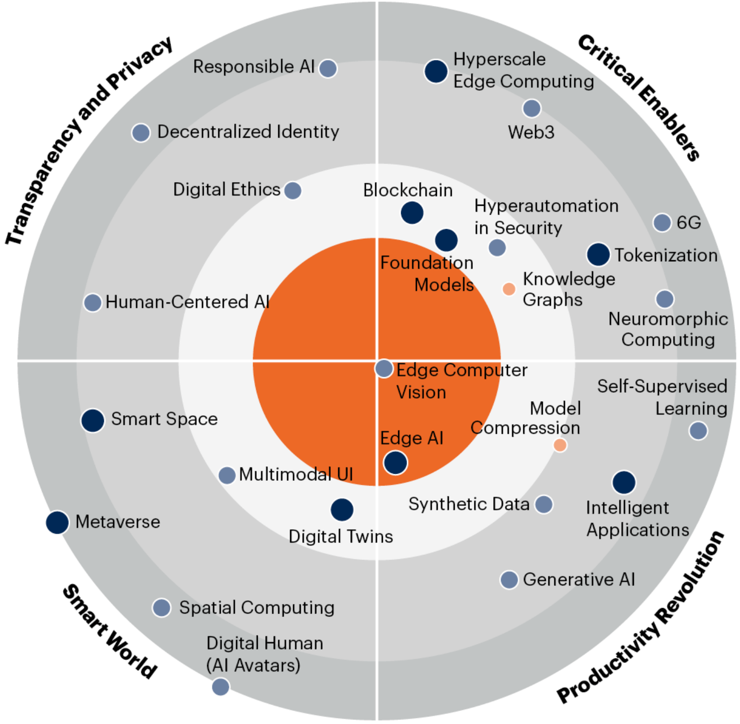
Doing something, does not mean doing it efficiently!



"I choose a lazy person to do a hard job. Because a lazy person will find an easy way to do it"
– *(attributed to Bill Gates, although its origin maybe much longer in time…)*

However, with the right tools, we can move the world!





Vision → dominant source of information!

INTERESTING VISION FACTS:

- Two thirds of the brain electrical activity (2/3 billion firings /s) when eyes open.
- 50% of our neural tissue directly (or indirectly) related to vision

  Source: R. S. Fixot, Neuroanatomist, 1957
- More neurons dedicated to vision than all four senses combined
- Olfactory cortex losing ground to visual cortex
  (i.e. vision is "eating" our smell!)

  Source: John Medina, Brain Rules, 2015

Vision is the BRAIN activity & NOT the sensor!

57 YEARS OF COMPUTER VISION
(AND COUNTING)

Put enough processing power behind a digital camera and you've got "computer vision," the process by which machines can analyze the visual world. Since the advent of the transistor, systems that can do this have become cheaper, faster, and smaller. Here's a quick overview of the highs and lows in the technology's history.

1957 → The first computer scanner copies a 2-inch photo of the inventor Russell A. Kirsch's son.

1964 → Defense contractors Woody Bledsoe, Helen Chan Wolf, and Charles Bisson launch a facial-recognition system for an unnamed intelligence agency.

1976 → U.K. police invent a license-plate recognition system. The first major installation is in 1993, as a "ring of steel" around London to counteract IRA bombings.

1985 → The first autonomous land vehicle, made by Lockheed Martin, Carnegie Mellon, and others, uses video-based imaging to follow a road at three mph.

2004 → Mars rovers Spirit and Opportunity land on the Red Planet using computer vision to calculate distance and position on descent.

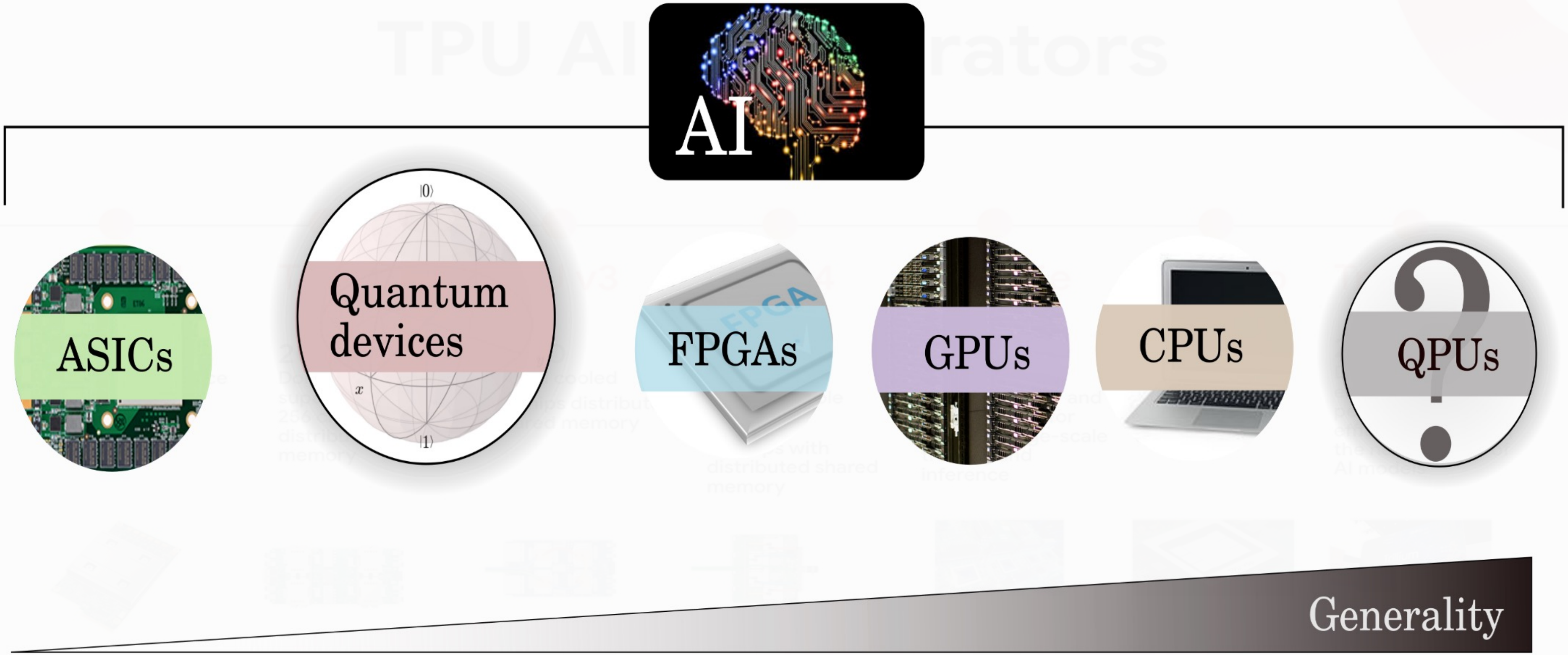
2008 → The first 3-D pizza-sorting system, the "Scorpion," builds a 3-D profile of 7,200 products per hour using multiple cameras. It automatically culls misshapen pies.

2010 → Microsoft Kinect is released; it can track 20 human features at 30 times per second. Shortly thereafter, a man hacks the device to track his own nipples for the first time.

2014 Phone processors become fast enough to handle pattern recognition. Apps such as Vhoto pick worthy stills from a video based on action sequences and facial expressions.

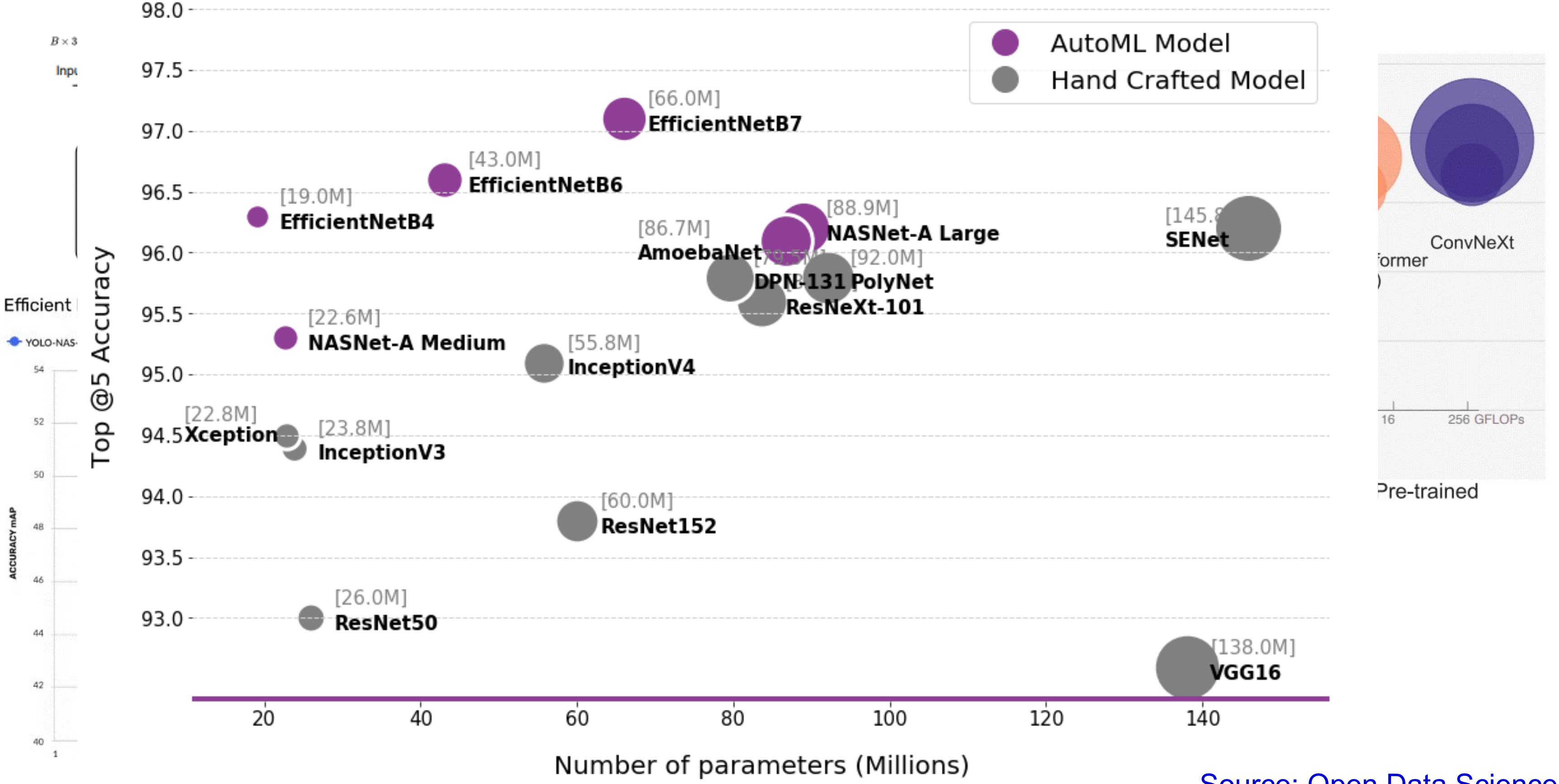
Source: Popular Science, July 2014

Architectures For the “Brain” computation...



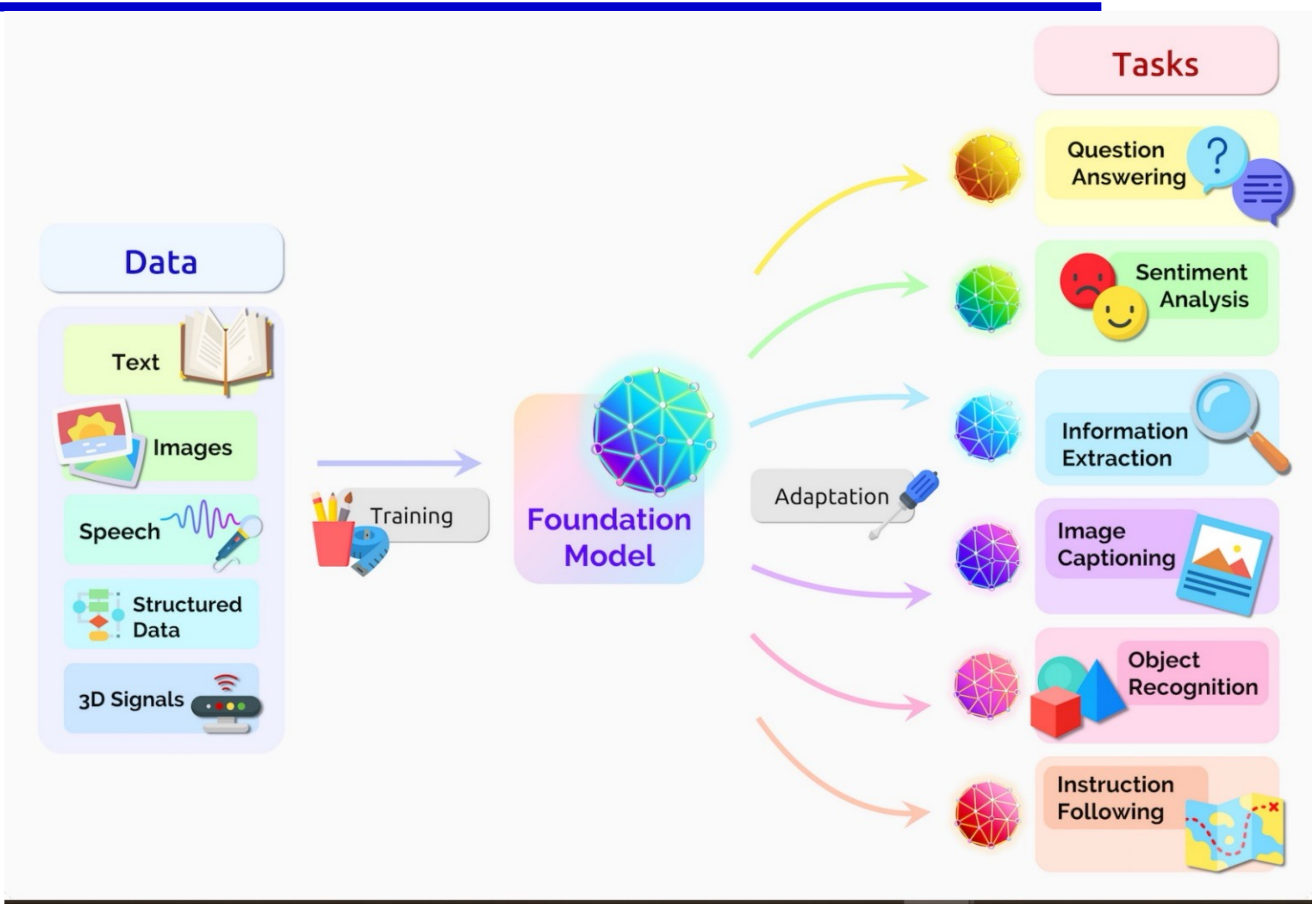
Source: Google, Tenstorrent

Deep Neural Networks are dominating!



Source: Open Data Science

Data→ Model → Adaptation

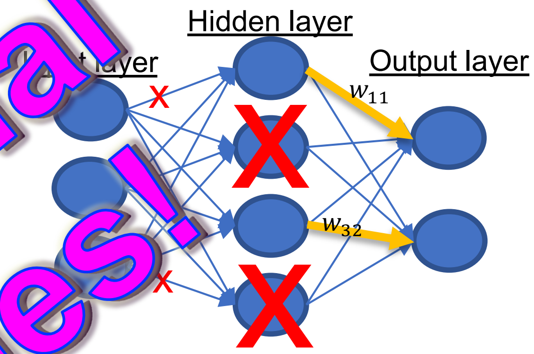


Ack: Nvidia

Energy and Performance-Efficient DNNs

- Software & EDA
  - Structured Pruning
  - Unstructured Pruning
  - Weight sharing
  - Quantization
- Hardware Design
  - Fixed point representation
  - Limited fixed point precision
  - Systolic array based designs

Computational Approaches!



$w_{11} = w_{32}$

$w_{11} \in \{1, 0, -1\}$

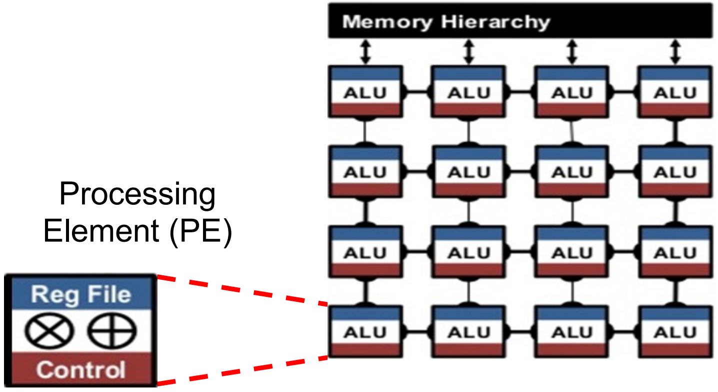
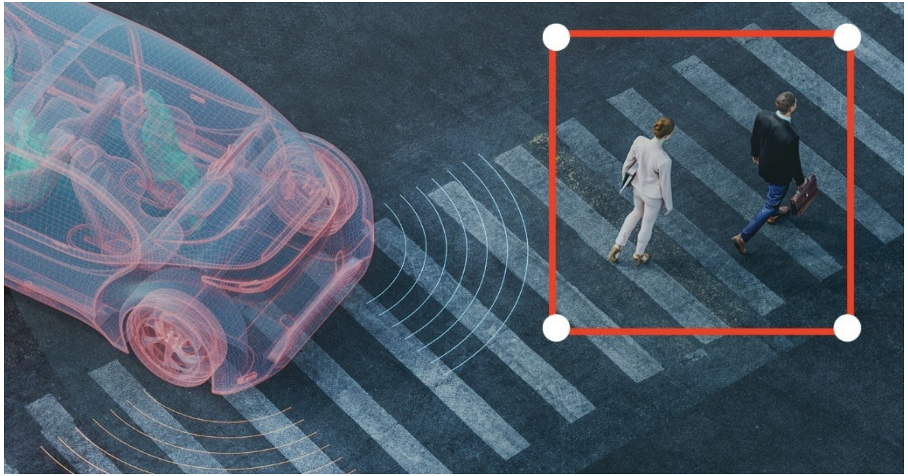


Image Source: Eyeriss@MIT

However, we are doing a lot!

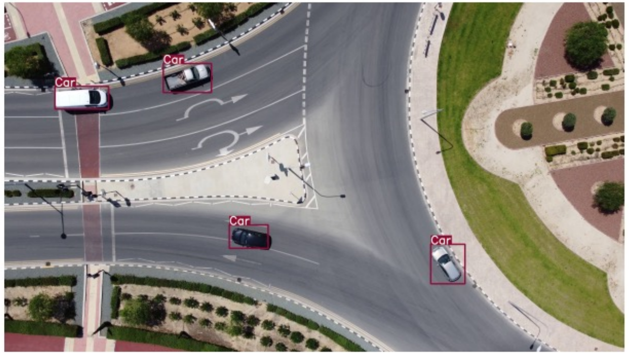
… in fact, way more than what is necessary!

- Aiming to design a model which performs good across all data!
- Benchmarking – ImageNet, CIFAR100, MNIST, etc.
- This is TOO MUCH!



I need the car to detect pedestrians.
I don't care if it's a man or a woman!

Let’s take another example!



(a) 50m

(b) 200m

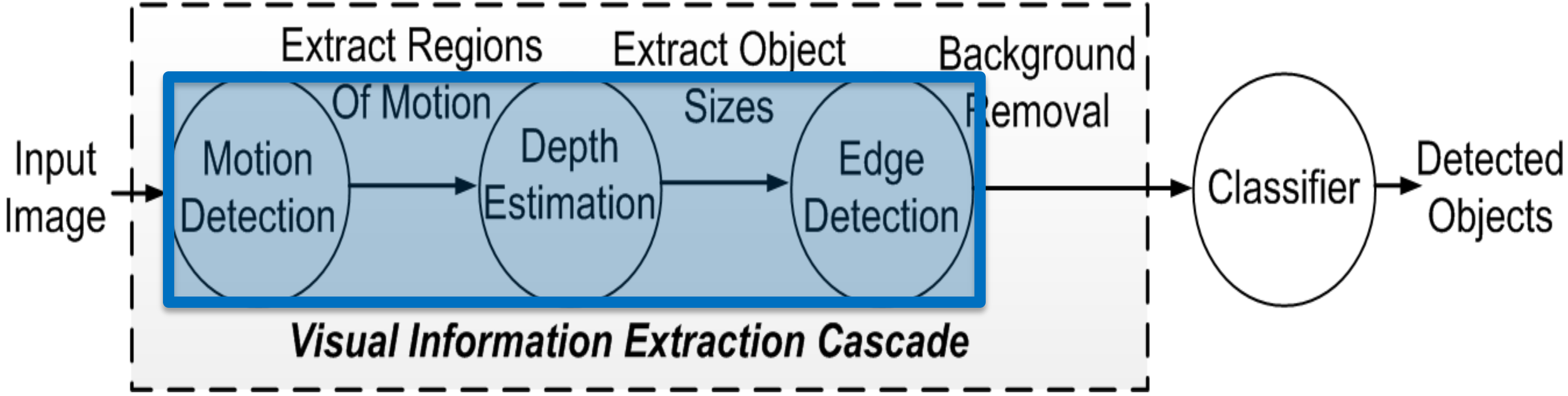
(c) 350m

(d) 400m

How do we (humans) do it?

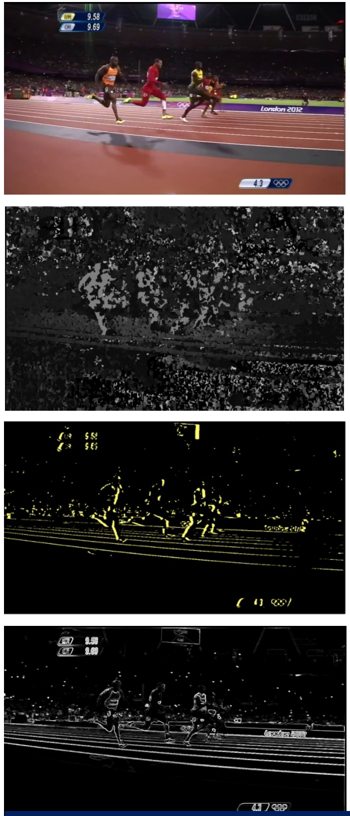
- **Multi-modal sensing**
  - I hear a sound, I turn and focus there!
- **Saliency (visual focus)**
  - cognitive ability to quickly differentiate from background
- **Short and Long Term Spatiotemporal Memory**
  - Focus on current context
  - Quick context switch
- **Multi-Task**
  - Most tasks done “mechanically” – i.e. set-it-and-forget-it.
  - Event-Driven Thinking!
- **Occasional Refresh**

Don't search everywhere! - Visual Information Extraction Cascade



- Step 1: motion (focus on changes)
- Step 2: depth (focus on object size)
- Step 3: edge (focus only where's a lot of information)
- Step 4: Classify *only what's necessary!*

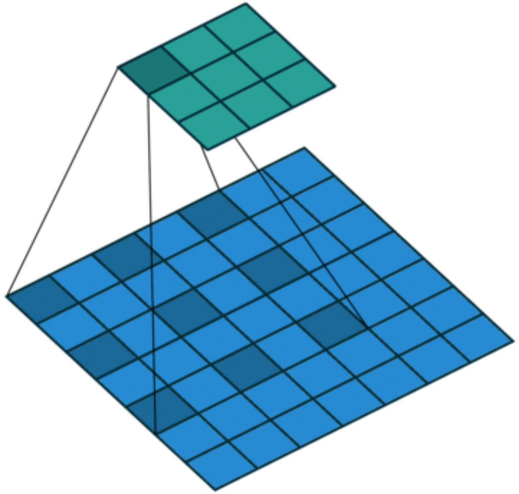
Process only what's necessary!

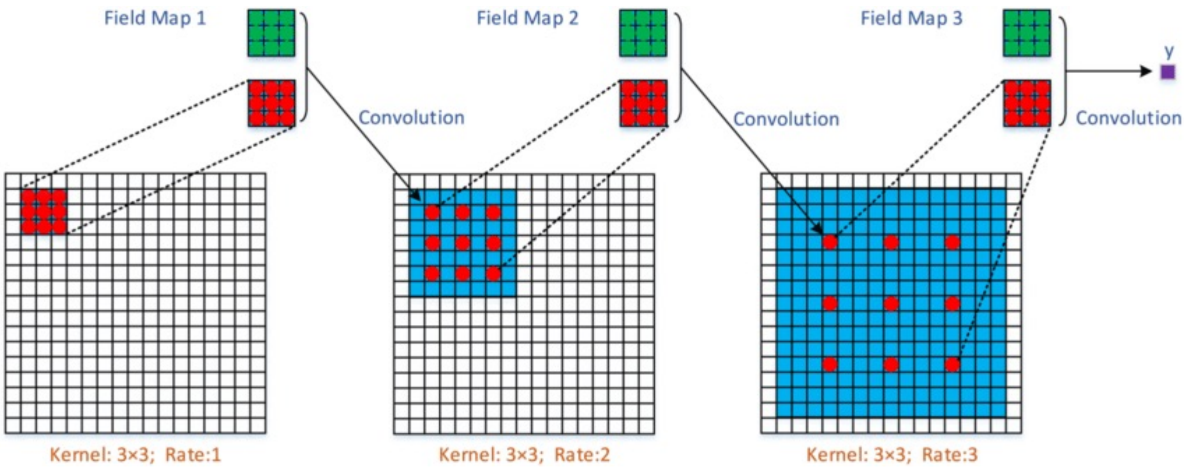


- A (very) high-speed 100m race!
- Cluttered background
- Various illumination conditions
- 20-76% overall reduction from motion
- 12-30% reduction from depth
- 15-20% reduction from edge
- **~1% of useful data reach the classifier**

Kyrkou, C., Theocharides, T., Bouganis, CS. et al. Embedded hardware-efficient real-time classification with cascade support vector machines. IEEE Transactions on Neural Networks and Learning Systems, Vol 27, Issue 1, pp. 99-112.

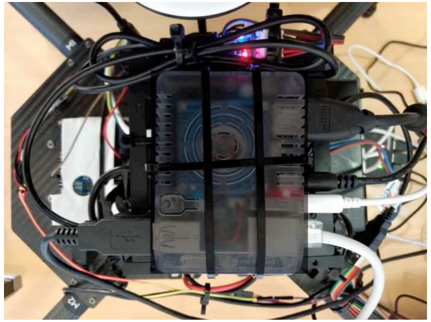
Another example –Atrus (dilated) Convolution

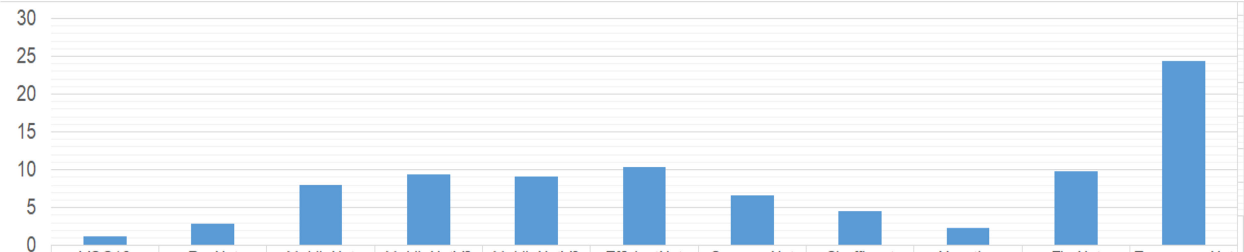




C. Kyrkou and T. Theocharides, "EmergencyNet: Efficient Aerial Image Classification for Drone-Based Emergency Monitoring Using Atrous Convolutional Feature Fusion," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1687-1699, 2020

Encouraging results!





	VGG16	ResNet	MobileNet	MobileNetV2	MobileNetV3	EfficientNet	SqueezeNet	Shufflenet	Xception	FireNet	EmergencyNet
FPS	1.147315282	2.906976744	7.987220447	9.360666479	9.090909091	10.41666667	6.618133686	4.566210046	2.314814815	9.854158455	24.3902439

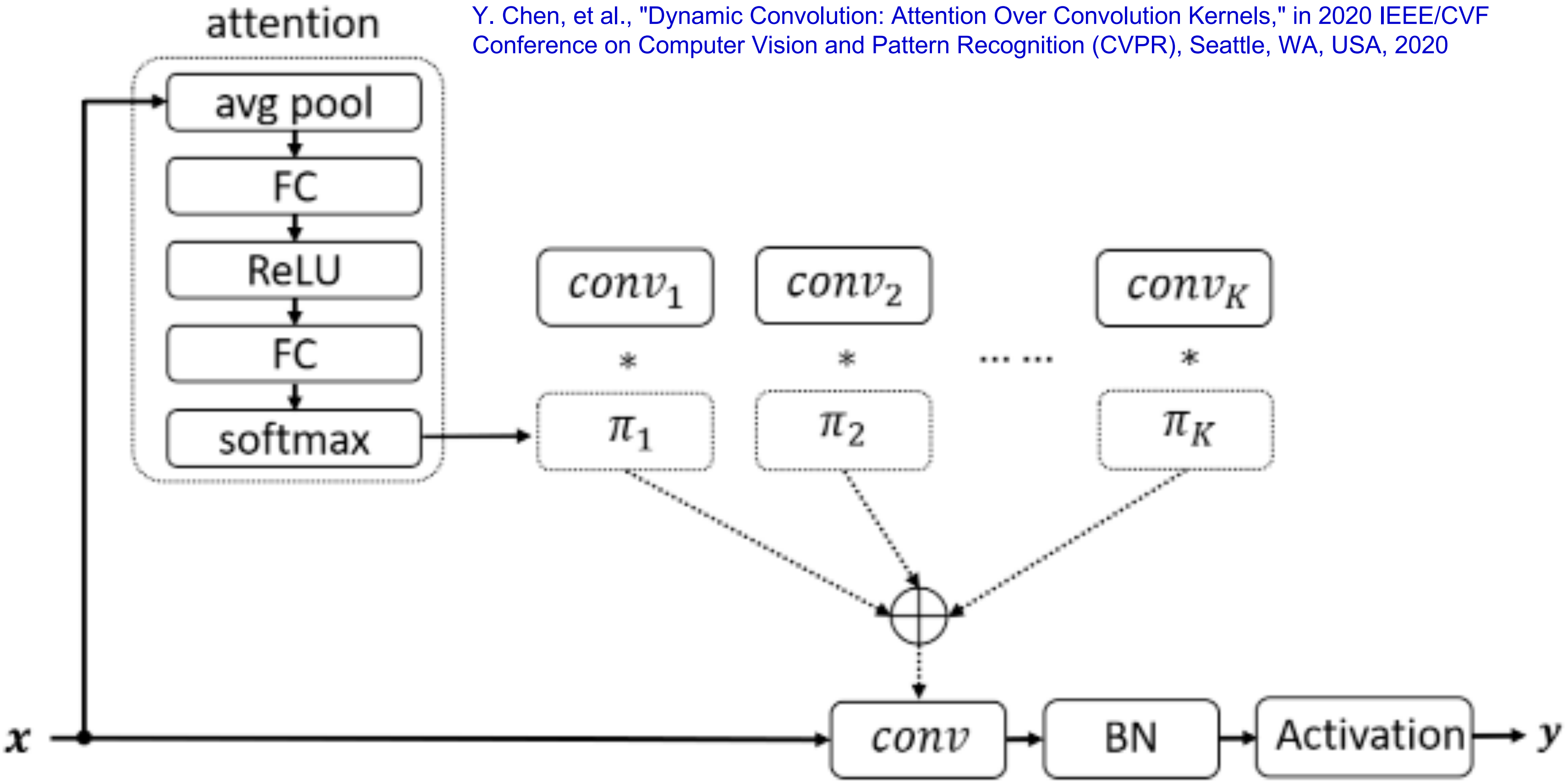
C. Kyrkou and T. Theocharides, "EmergencyNet: Efficient Aerial Image Classification for Drone-Based Emergency Monitoring Using Atrous Convolutional Feature Fusion," in *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 13, pp. 1687-1699, 2020, doi: 10.1109/JSTARS.2020.2969809.

→ Dynamic Deep Neural Networks!

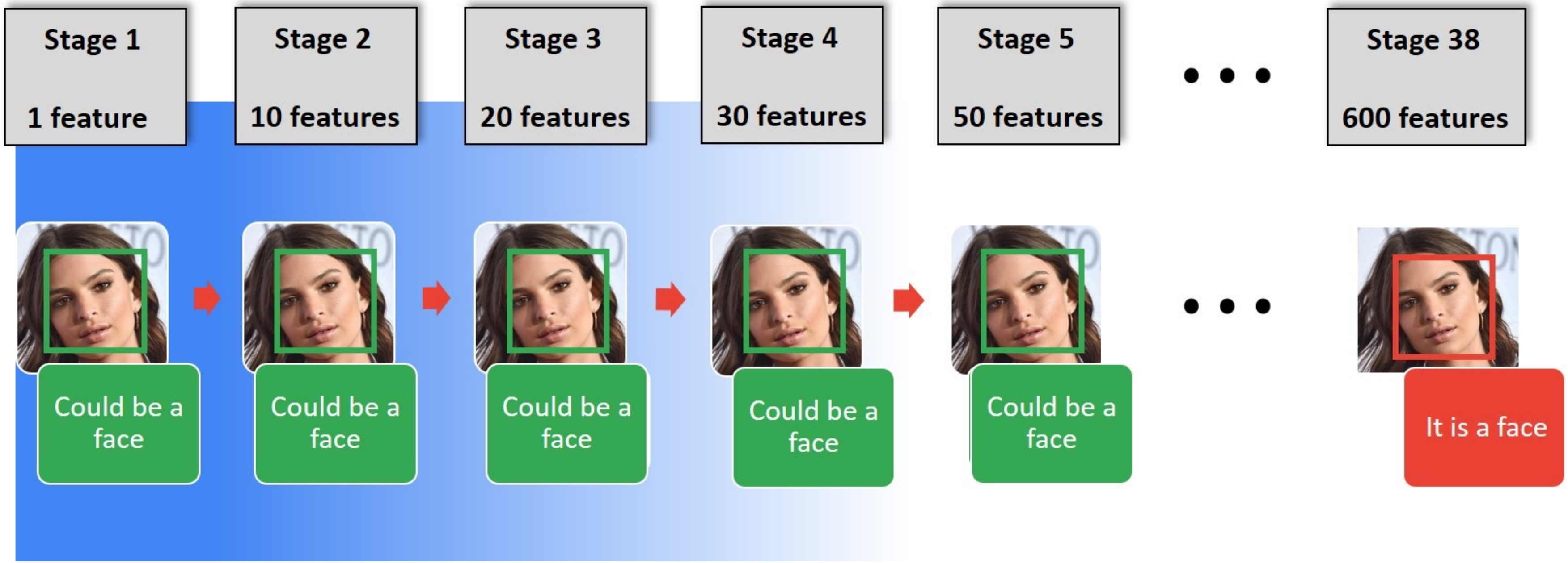
- Improved Efficiency: Computation and energy reduction by *adapting the model's depth or parameters to each input*.
- Faster Inference: Early exit strategies for example allow *quick predictions* for easy inputs, lowering latency.
- Resource Awareness: Dynamic models can *adjust operations* to handle resource constraints
- Maintained Accuracy: Often preserve accuracy by allocating full model capacity *only when needed*.

Example: Dynamic Convolution

Y. Chen, et al., "Dynamic Convolution: Attention Over Convolution Kernels," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020



Early Exit Deep Neural Networks



Throwback: Adaboost (Viola & Jones, 2001)

Using FF training, *focus on optimizing each layer!*

Layer-wise Loss Function with Channel-wise Competitive Learning

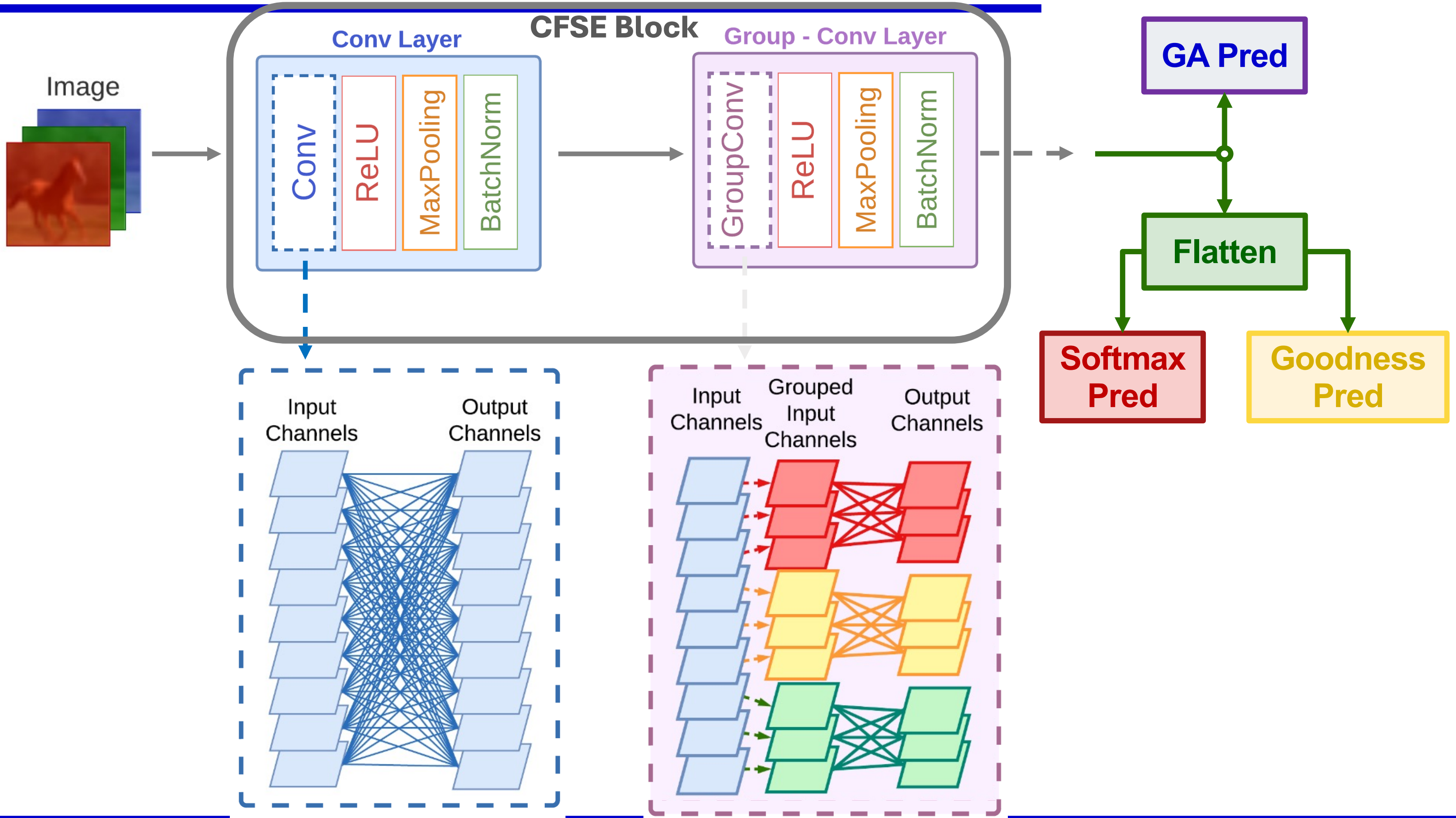
- Reformulates the goodness function to avoid negative data construction
- Enables each CNN layer to act as an independent classifier

Channel-wise Feature Separator and Extractor (CFSE) Block

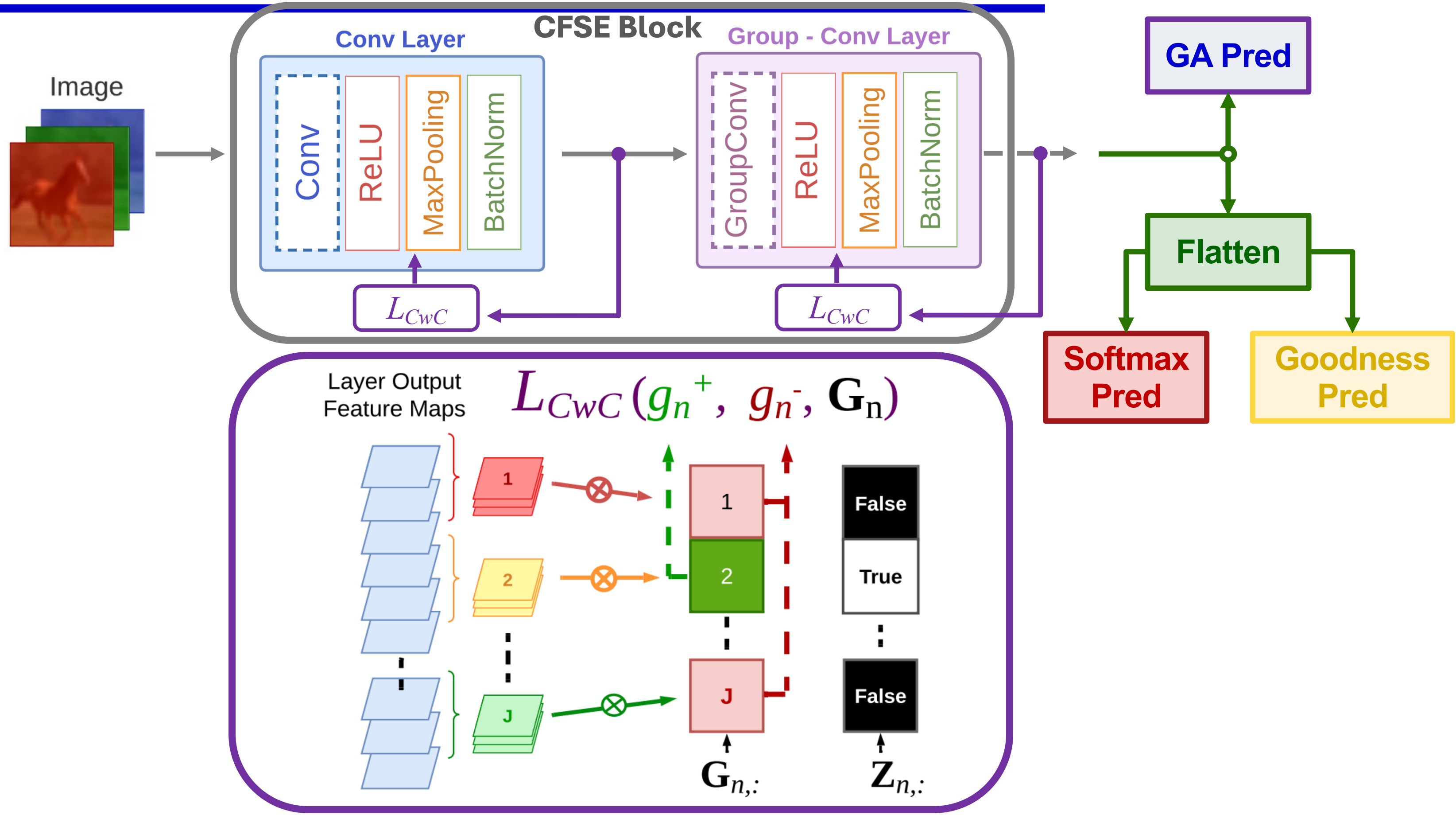
- Incorporates channel-wise grouped convolutional layers
  - Partitions feature space
- Facilitates learning of compositional features via standard non-separable convolutional layers

A. Papachristodoulou, C. Kyrkou, S. Timotheou & T. Theocharides, "Convolutional Channel-Wise Competitive Learning for the Forward-Forward Algorithm", Proceedings of the AAAI Conference on Artificial Intelligence, 2024

Competitive Layer-Wise Learning Model Architecture

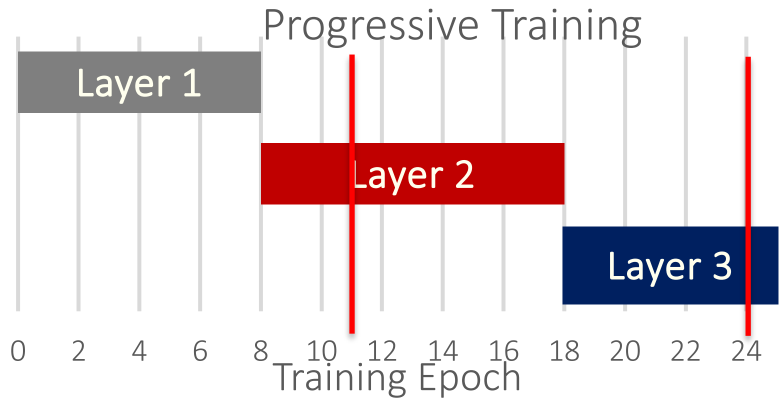


Model Architecture

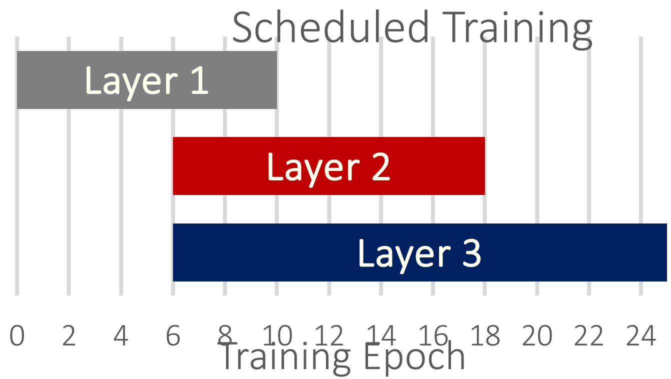


Benefits of layer-wise learning

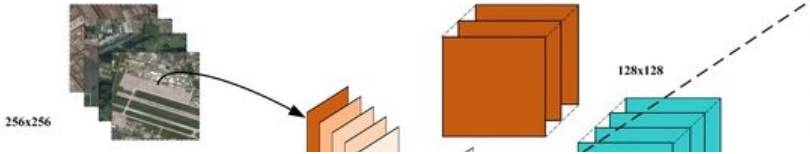
Progressive Training

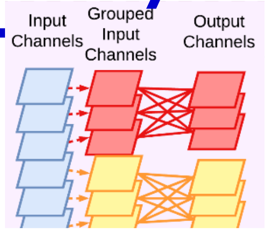


Inter-Leaved Layer (ILT) Training

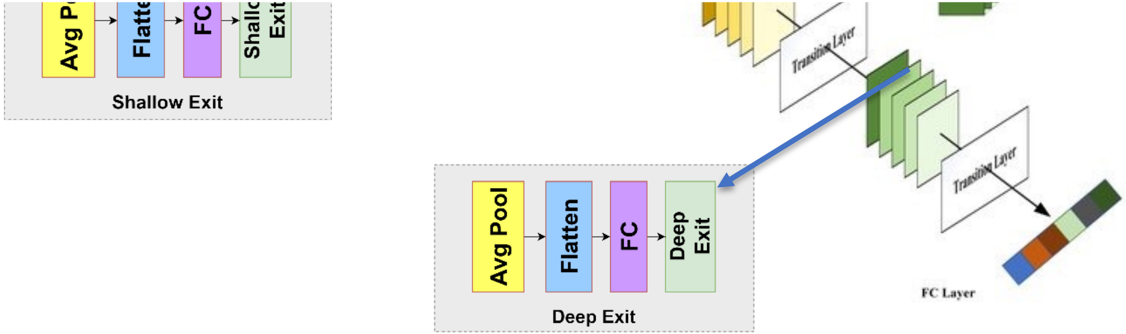


Allows for Promising Performance w/ Early Exits!

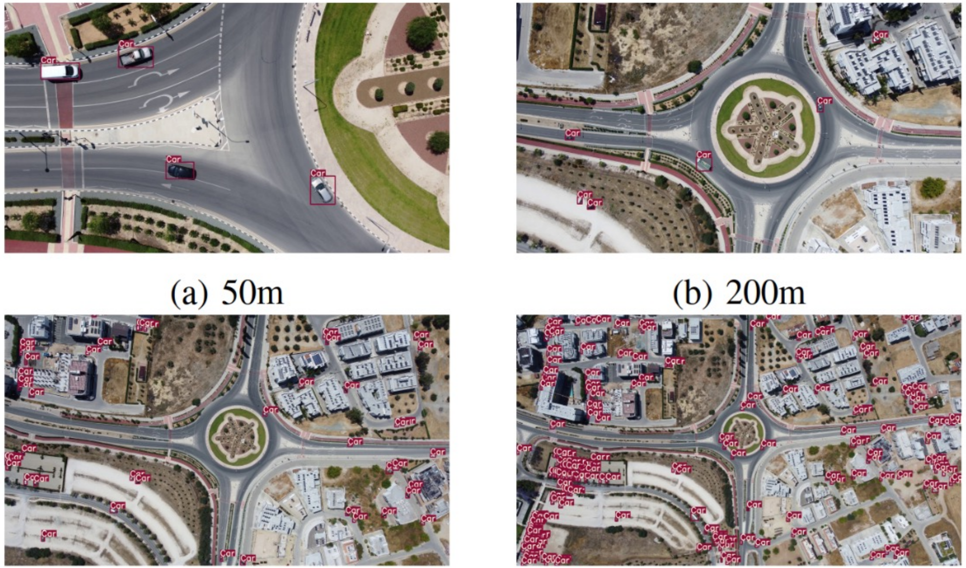




Performance				
Model	Accuracy	MAC Operations	Energy Consumption / Sample	Latency / Sample
End-to-End	85.4%	100%	100%	100%
Deep Exit	85.33%	92.08%	77.28%	80.45%
Shallow Exit	85.13%	76.54%	62.40%	65.50%



Case-Study: UAV Object Detection



(a) 50m (b) 200m (c) 350m (d) 400m

MAC Operations (M)

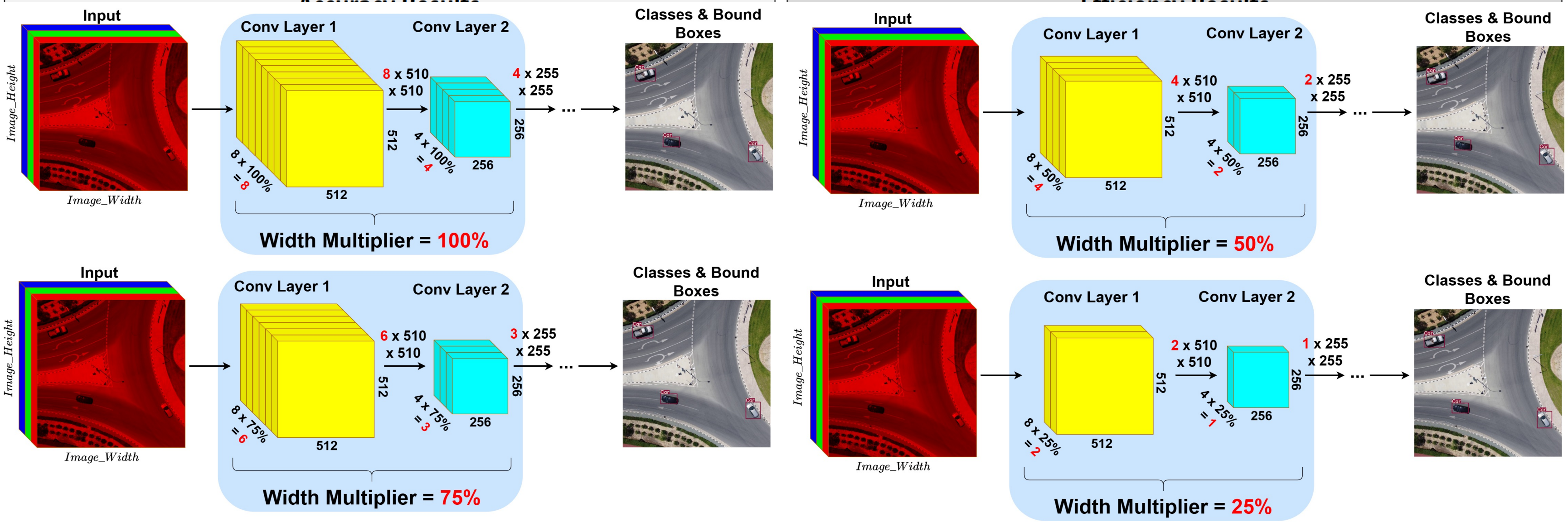
Resolution	1	0.75	0.5	0.25	0.1	0.01
1536	89,03	66,9	44,77	22,65	11,38	4,67
1408	75,92	57,05	38,18	19,31	9,71	3,98
1280	61,19	45,98	30,77	15,57	7,83	3,21
1088	45,42	34,13	22,84	11,55	5,81	2,38
896	31,99	24,04	16,09	8,14	4,09	1,68
768	22,71	17,07	11,42	5,78	2,9	1,19
640	16,35	12,29	8,22	4,16	2,09	0,86
512	11,03	8,29	5,55	2,81	1,41	0,58
384	6,75	5,07	3,39	1,72	0,86	0,35
256	2,92	2,19	1,47	0,74	0,37	0,15

Size (MB)

Resolution	1	0.75	0.5	0.25	0.1	0.01
1536	22,55	17,29	13,53	11,27	10,69	10,54
1408	21,01	15,75	11,99	9,73	9,15	9
1280	19,27	14,01	10,25	7,99	7,41	7,26
1088	17,41	12,15	8,39	6,13	5,55	5,4
896	15,83	10,57	6,81	4,55	3,97	3,82
768	14,73	9,47	5,71	3,45	2,87	2,72
640	13,98	8,72	4,96	2,7	2,12	1,97
512	13,35	8,09	4,33	2,07	1,49	1,34
384	12,85	7,59	3,83	1,57	0,99	0,84
256	12,4	7,14	3,38	1,12	0,54	0,39

Width Multiplier

Experimenting with various Deep CNNs



512	0,938	0,642	0,499	0,465	0,384	0,0185
384	0,483	0,4	0,294	0,00335	0,00038	0,00037
256	0,00547	0,133	7,6E-05	8E-05	4,6E-05	7,1E-05

Width Multiplier

512	0,802	0,318	0,26	0,0591	0,0569	0,0701
384	0,0211	0,274	0,0324	0,00188	0,00017	0,0308
256	0,0154	0,0174	0,0157	0,0001	0,0025	0,00687

Width Multiplier

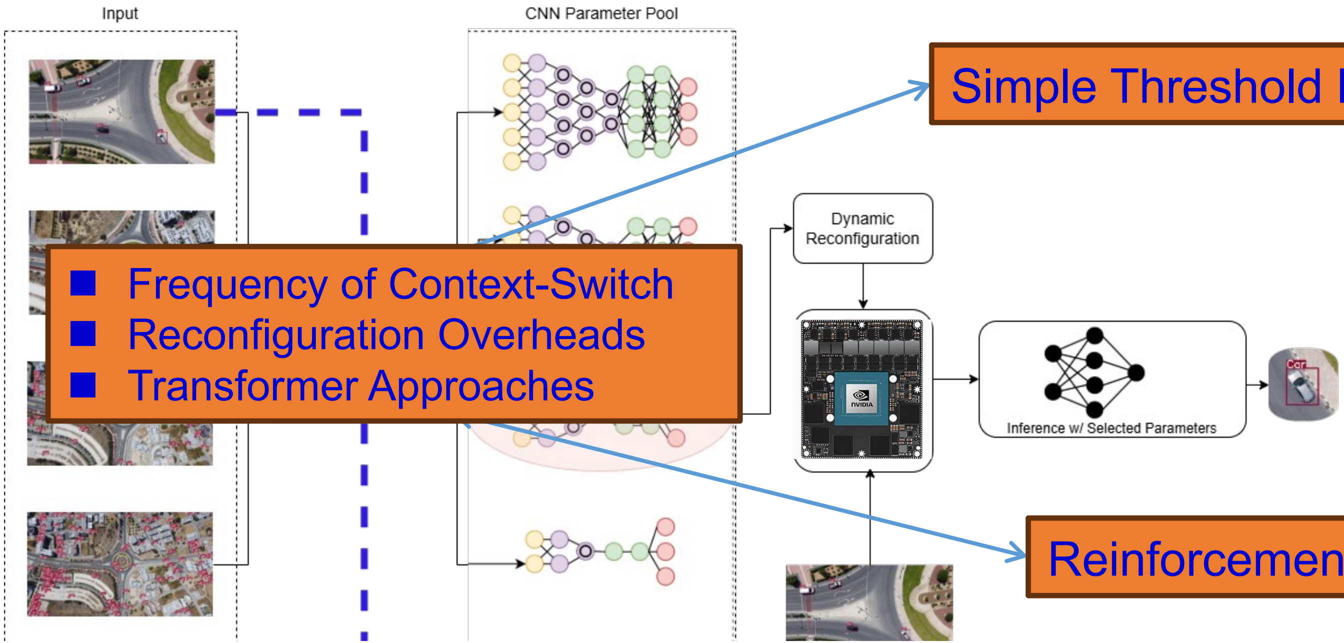
512	374,44	368,99	325,77	306,53	300,9	301,16
384	345,77	319,44	295,32	286,01	281,45	281,94
256	312,11	294,87	286,08	280,55	274,89	280,45

Width Multiplier

512	0,98	0,81	0,66	0,39	0,34	0,32
384	0,92	0,72	0,54	0,35	0,33	0,3
256	0,68	0,57	0,52	0,33	0,32	0,27

Width Multiplier

Context-Aware Dynamic Convolution Selector



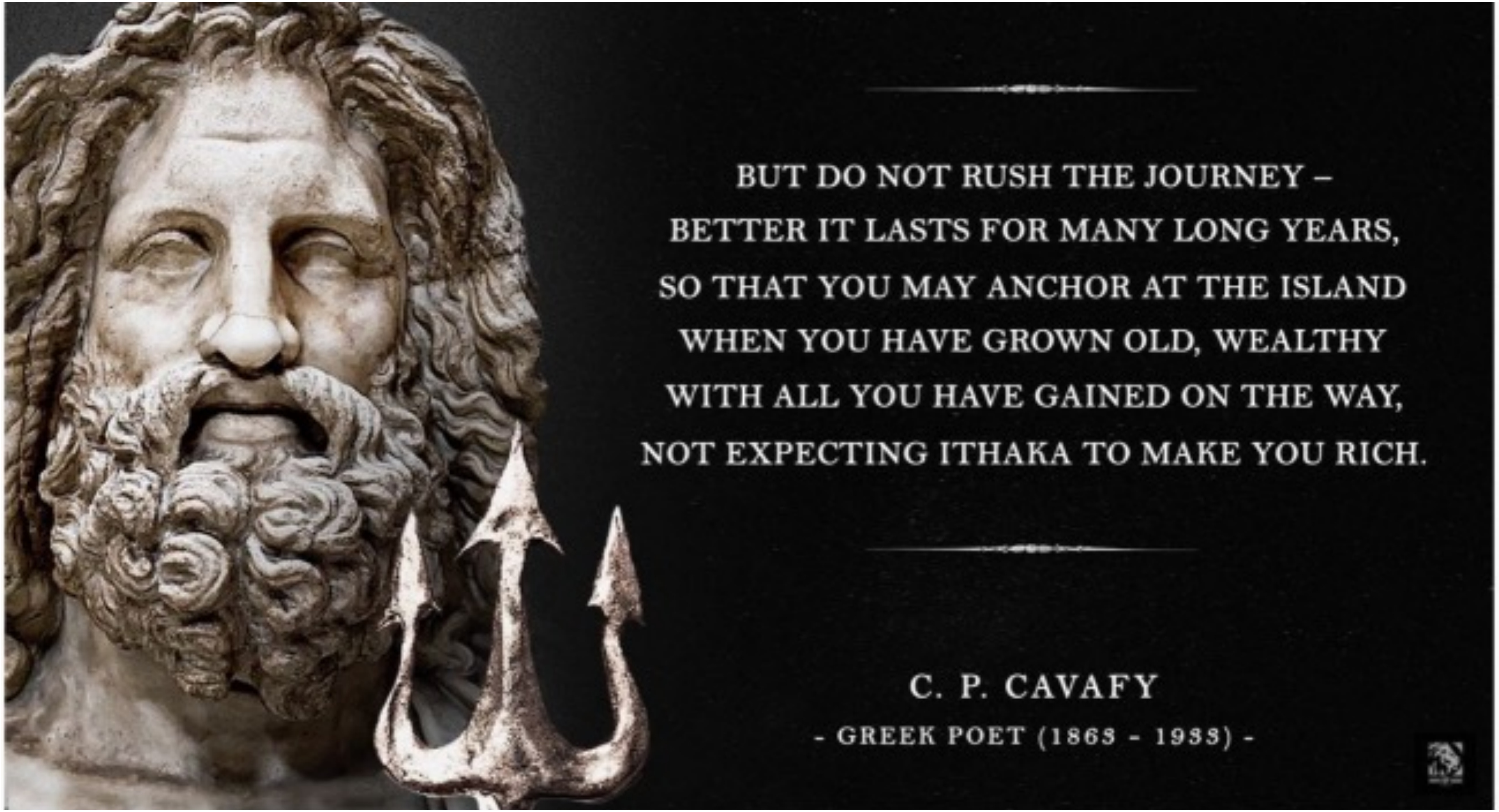
- Frequency of Context-Switch
- Reconfiguration Overheads
- Transformer Approaches

Simple Threshold Mechanism

Reinforcement Learning?

Model Configuration	Accuracy (mAP@.5)	CPU Usage (%)	Memory Usage (MB)	Power Usage (W)	Latency (ms)
Static	91.9	64.23	507.99	1.02	976.1
High Altitudes	92.6	52.68	429.36	0.75	805.6
Low Altitudes	93.5	43.51	359.08	0.56	373.23

Remember – It's not the destination, it's HOW to get there.



Thank you!

Acknowledgements:
Students, Post-Docs, Colleagues, Friends and Collaborators ☺

Parts of this work were supported by the European Union (i. Horizon 2020 Teaming, KIOS CoE, No. 739551, and from the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation, and Digital Policy.
Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Funded by: