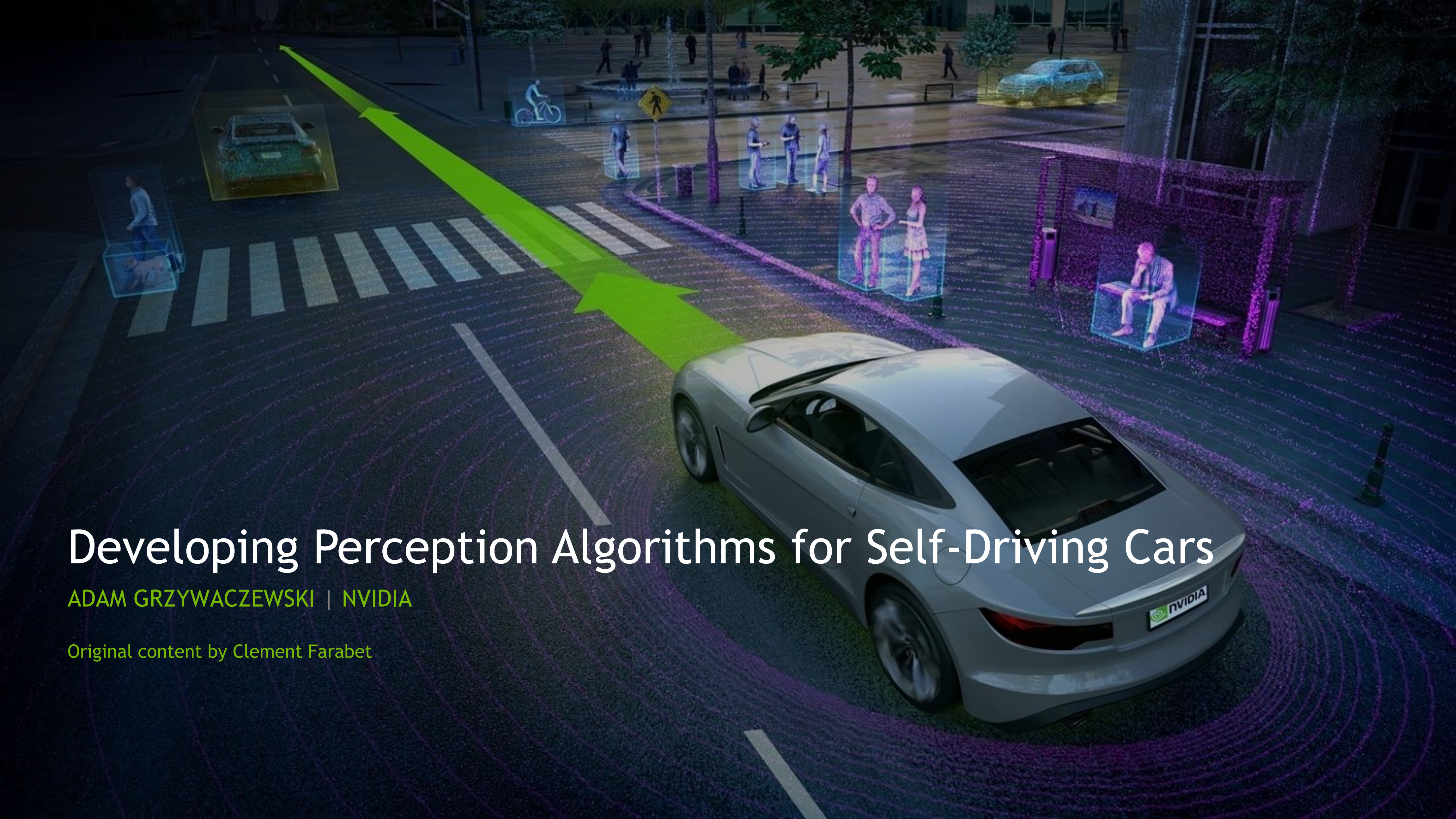


Developing Perception Algorithms for Self-Driving Cars

ADAM GRZYWACZEWSKI | NVIDIA

Original content by Clement Farabet

ABOUT ME

Adam Grzywaczewski – adamg@nvidia.com

- Deep Learning Solution Architect @ NVIDIA - Supporting delivery of AI / Deep Learning solutions
- 10 years expereince deliverying Machine Learning of all scale (from embedded, mobile to Big Data)
- My past experience:
  - Capgemini: https://goo.gl/MzgGbq
  - Jaguar Land Rover Research: https://goo.gl/ar7LuU

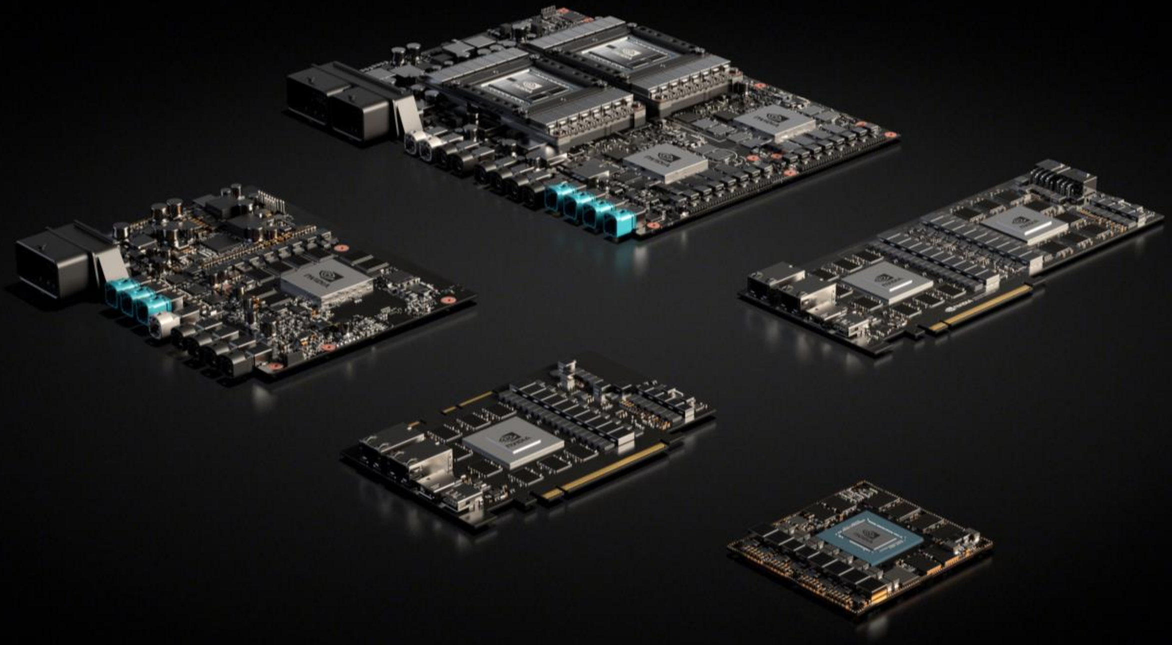
CONTEXT


summit
IBM

NVIDIA AGX

EMBEDDED AI HPC

High-speed SerDes — 109 Gbps + 320 Gbps I/O
Up to 320 TOPS Tensor Ops
Up to 25 TFLOPS FP32
Up to 16 GIGA Rays
Starting from 15W

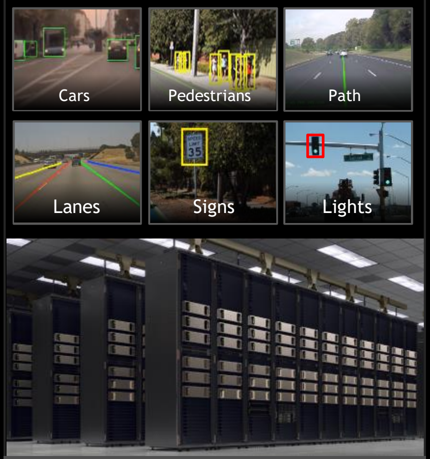


End to End System for AV

COLLECT DATA

1 PB+ per car per year

TRAIN MODELS



10+ DNNs
1M+ images per DNN

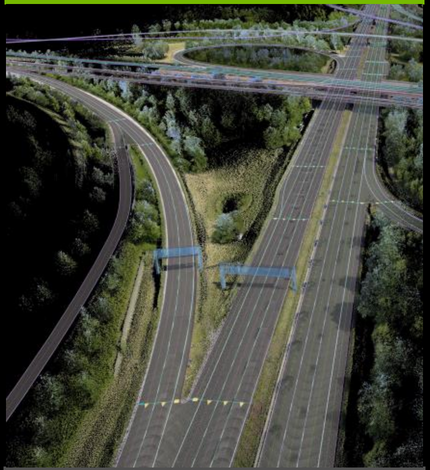
SIMULATE

10B+ miles to
ensure safe driving

RESIMULATE

Validate and verify
for self-driving cars
by 2020

MAPPING

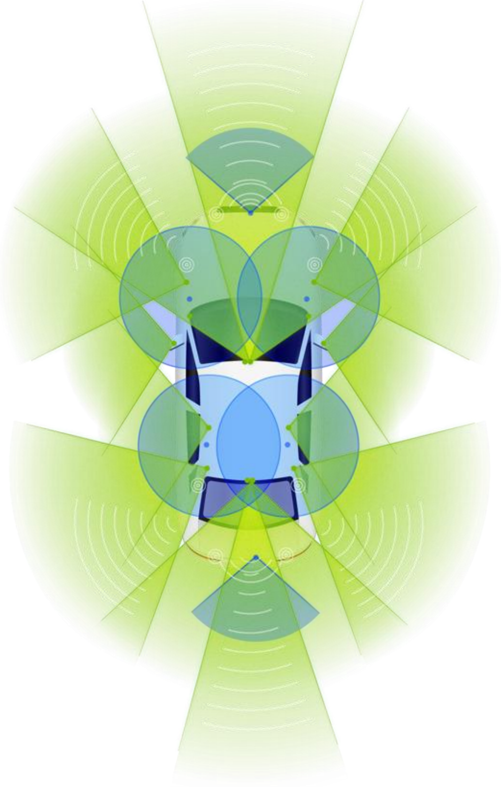


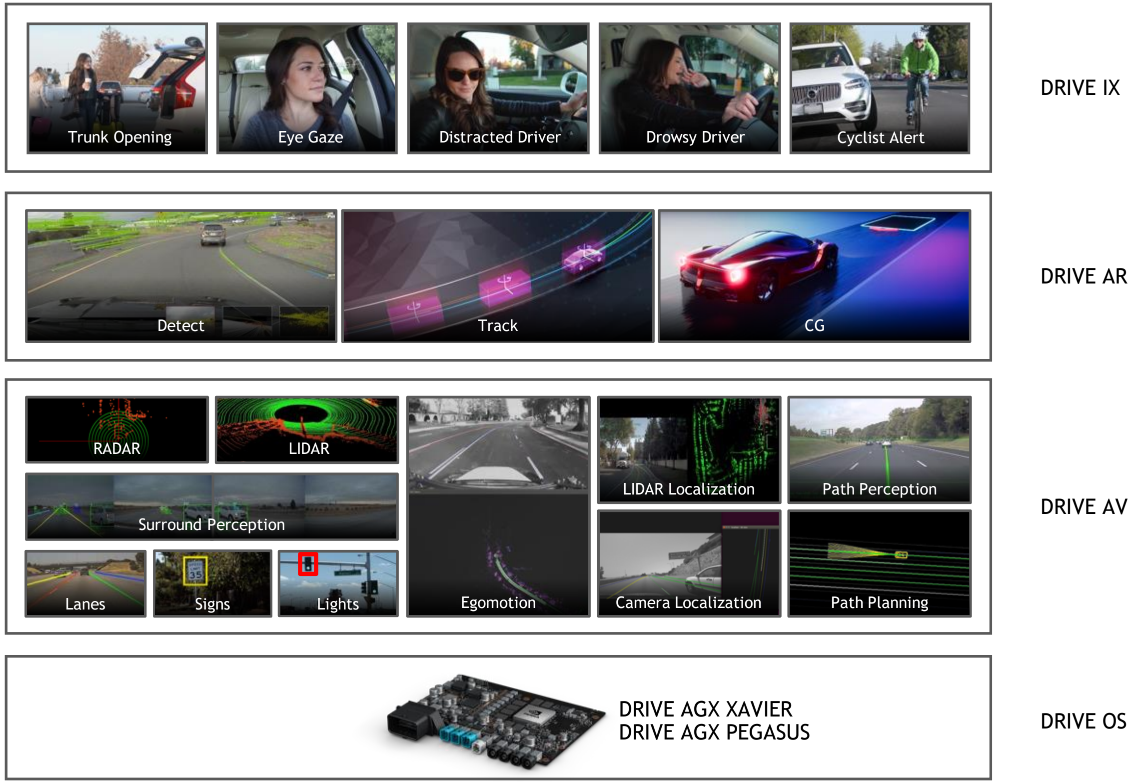
5+ DNNs
Create and update
HD maps

Sources: NVIDIA, RAND Corporation

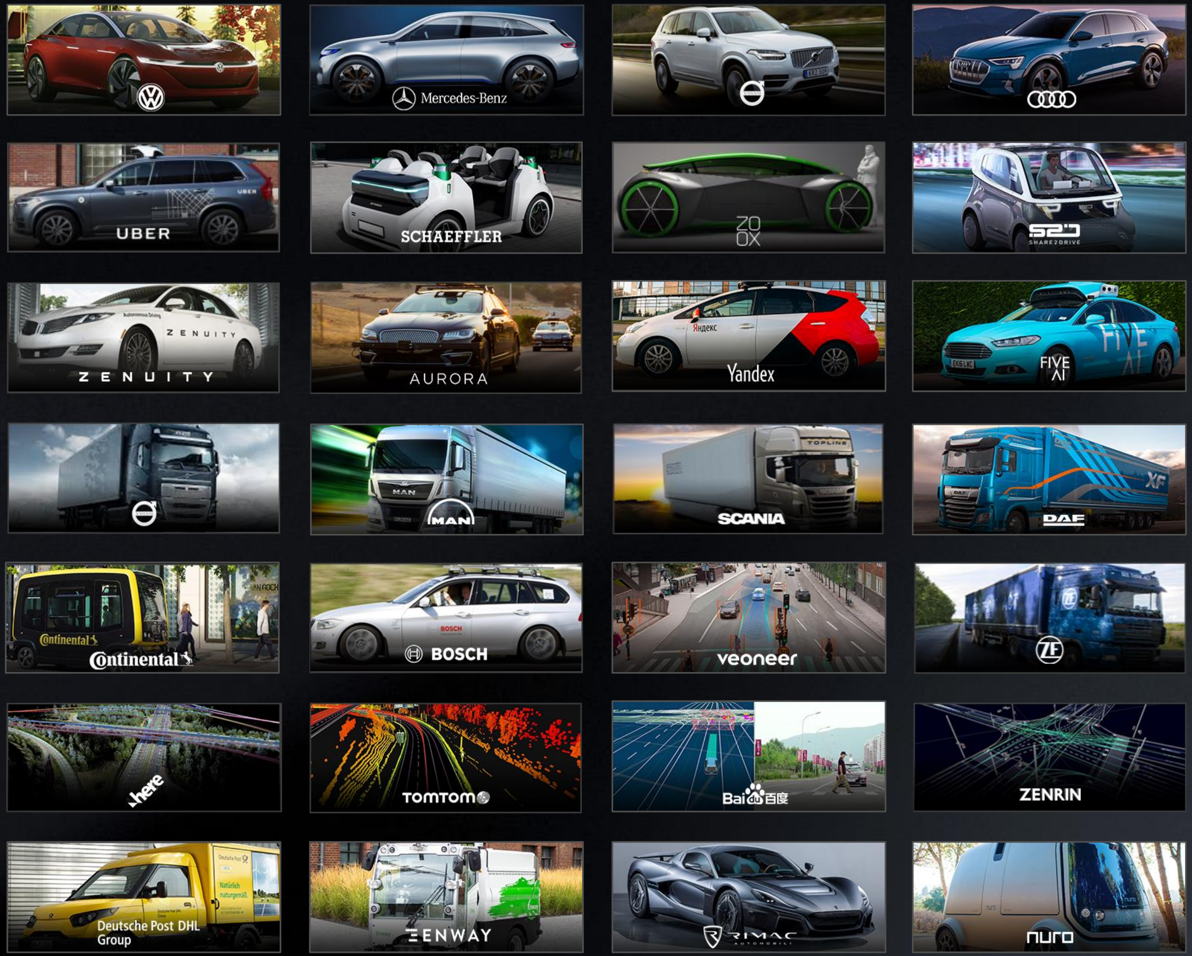
NVIDIA DRIVE: SOFTWARE-DEFINED CAR

Powerful and Efficient AI, CV, AR, HPC | Rich Software Development Platform
Functional Safety | Open Platform | 370+ partners developing on DRIVE





NVIDIA DRIVE PLATFORM ADOPTION ACROSS TRANSPORTATION



CARS

MOBILITY SERVICES

TRUCKS

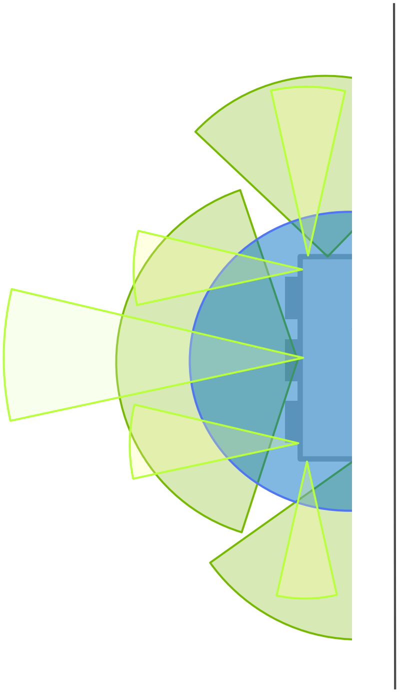
TIER ONES

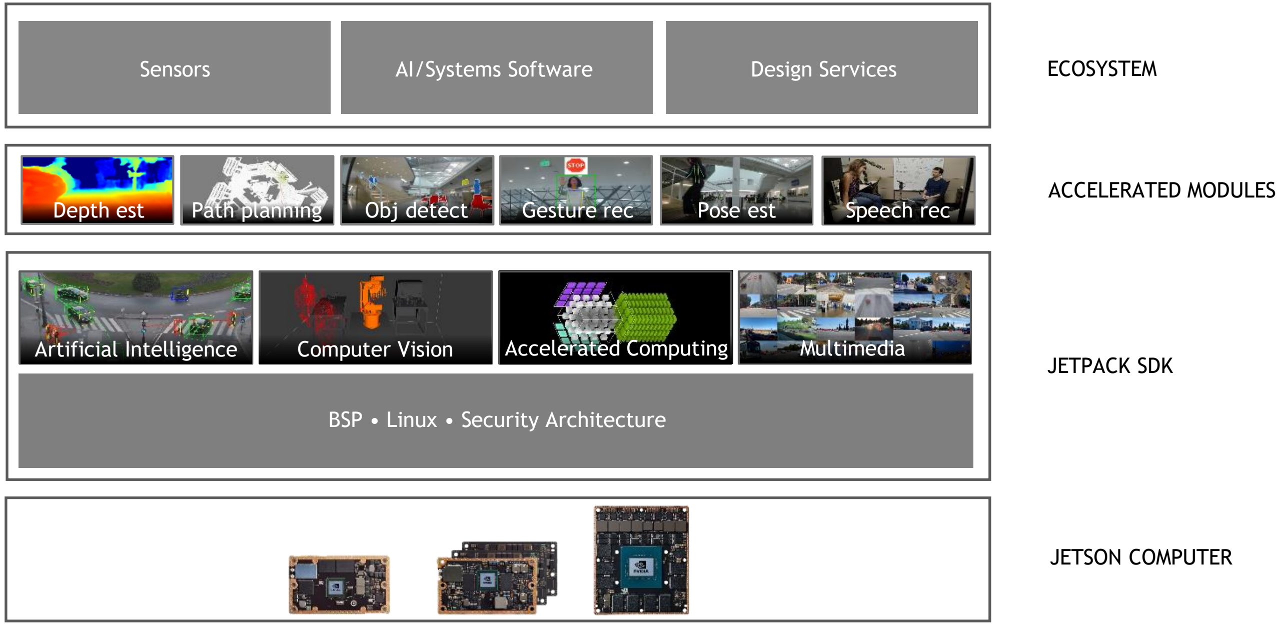
MAPPING

SPECIALTY

NVIDIA JETSON
SOFTWARE-DEFINED AUTONOMOUS MACHINES

Powerful and efficient AI, CV, HPC | Rich Software Development Platform
Open Platform | 200K Developers





BILLIONS OF AUTONOMOUS MACHINES

Taking advantage of the progress in self driving cars

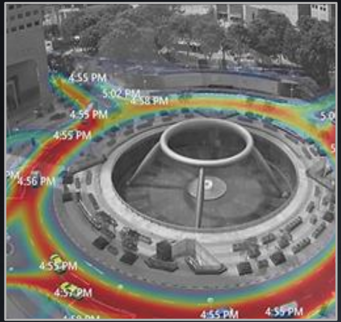
Industrial

Aerospace/Defense

Healthcare

Construction

Agriculture



Smart City

Retail

Logistics

Inventory Mgmt

Delivery

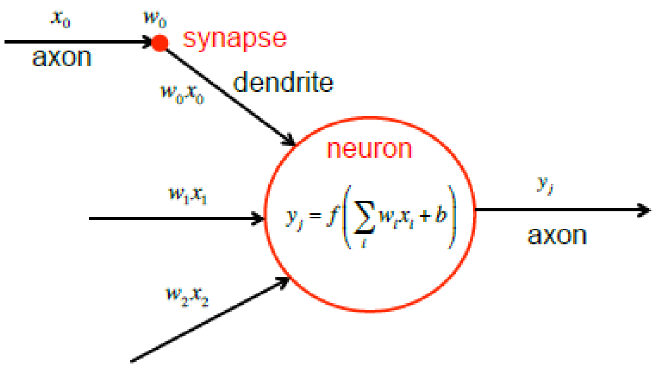
Inspection

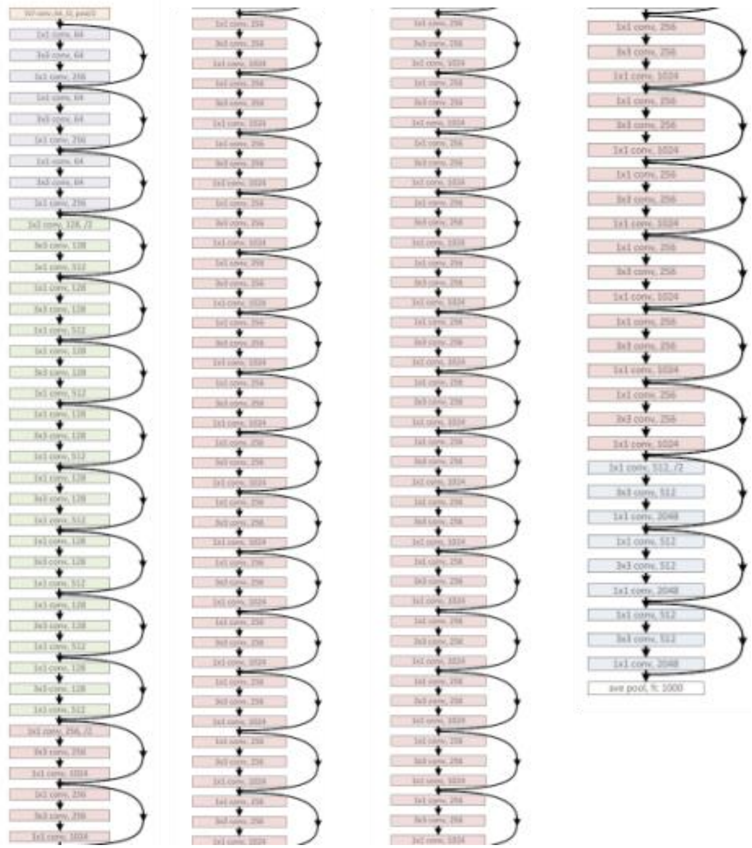
Service

DEEP LEARNING

NEURAL NETWORKS ARE NOT NEW

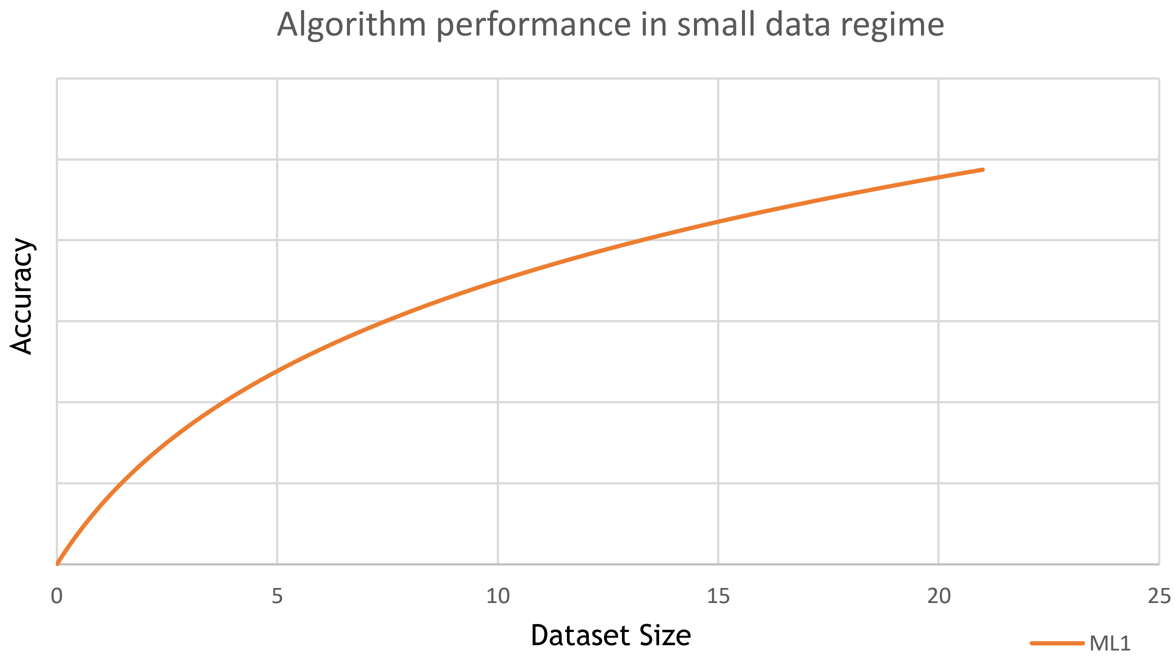
And are surprisingly simple as an algorithm





NEURAL NETWORKS ARE NOT NEW

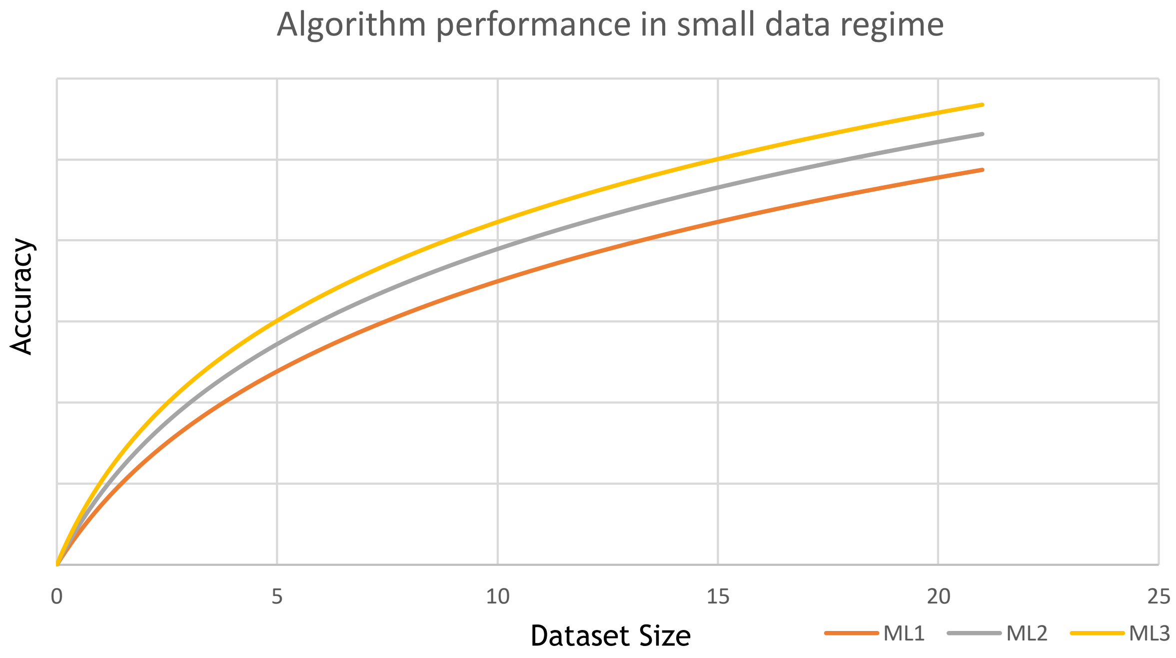
They just historically never worked well



Andrew Ng, “Nuts and Bolts of Applying Deep Learning”, https://www.youtube.com/watch?v=F1ka6a13S9I

NEURAL NETWORKS ARE NOT NEW

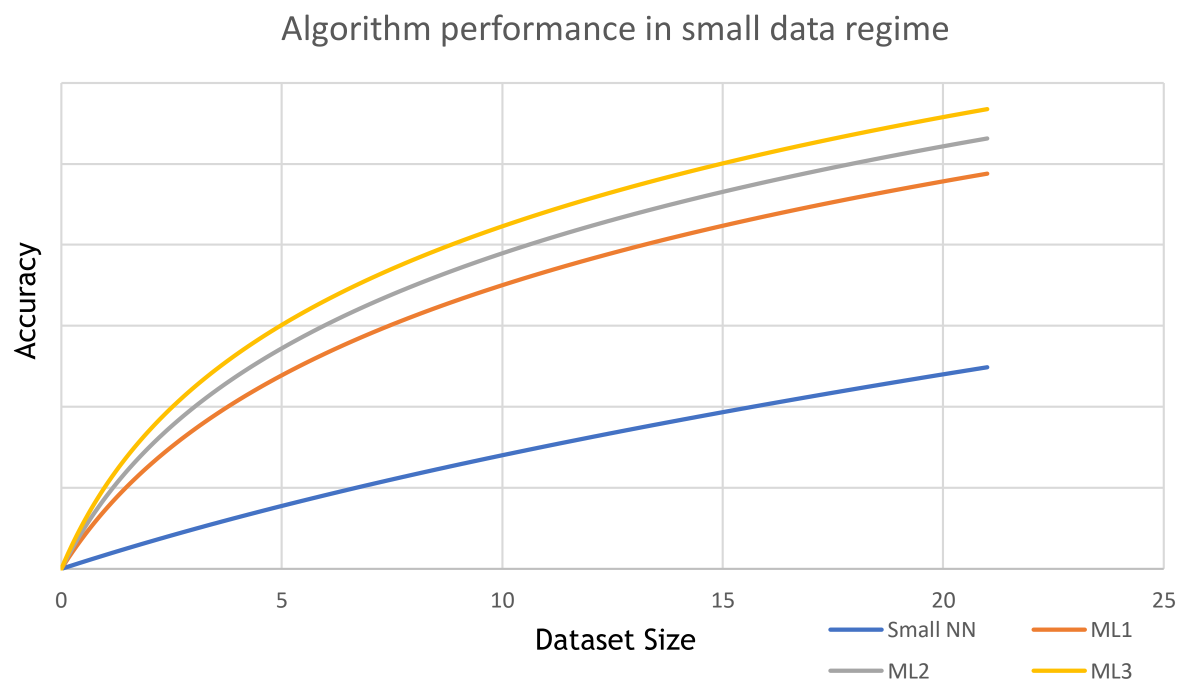
They just historically never worked well



Andrew Ng, “Nuts and Bolts of Applying Deep Learning”, https://www.youtube.com/watch?v=F1ka6a13S9I

NEURAL NETWORKS ARE NOT NEW

They just historically never worked well

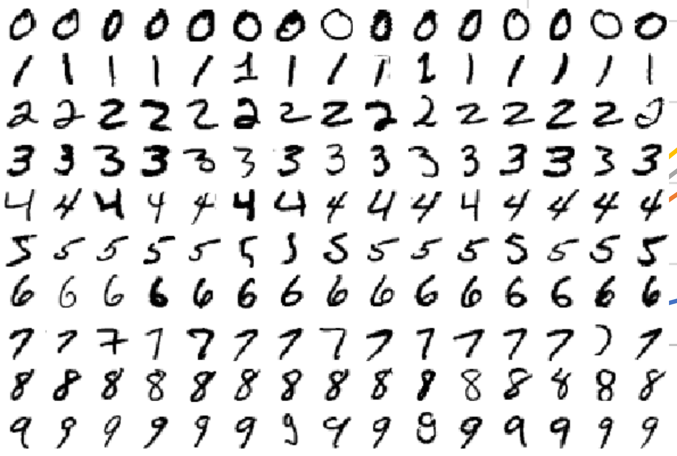


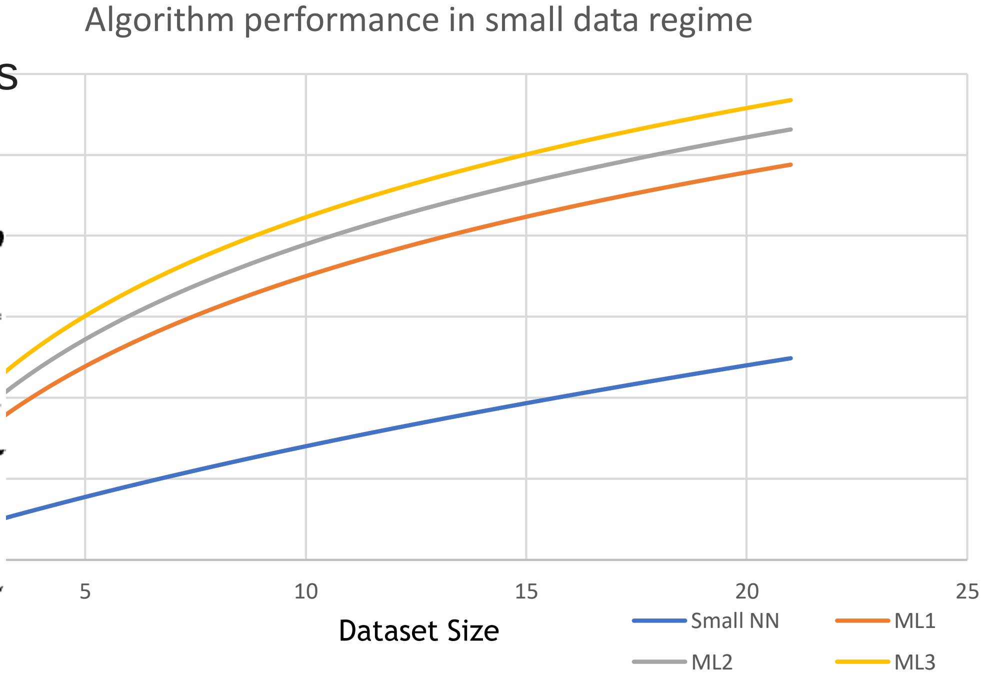
Andrew Ng, “Nuts and Bolts of Applying Deep Learning”, https://www.youtube.com/watch?v=F1ka6a13S9I

NEURAL NETWORKS ARE NOT NEW

Historically we never had large datasets or computers

The MNIST (1999) database contains 60,000 training images and 10,000 testing images.

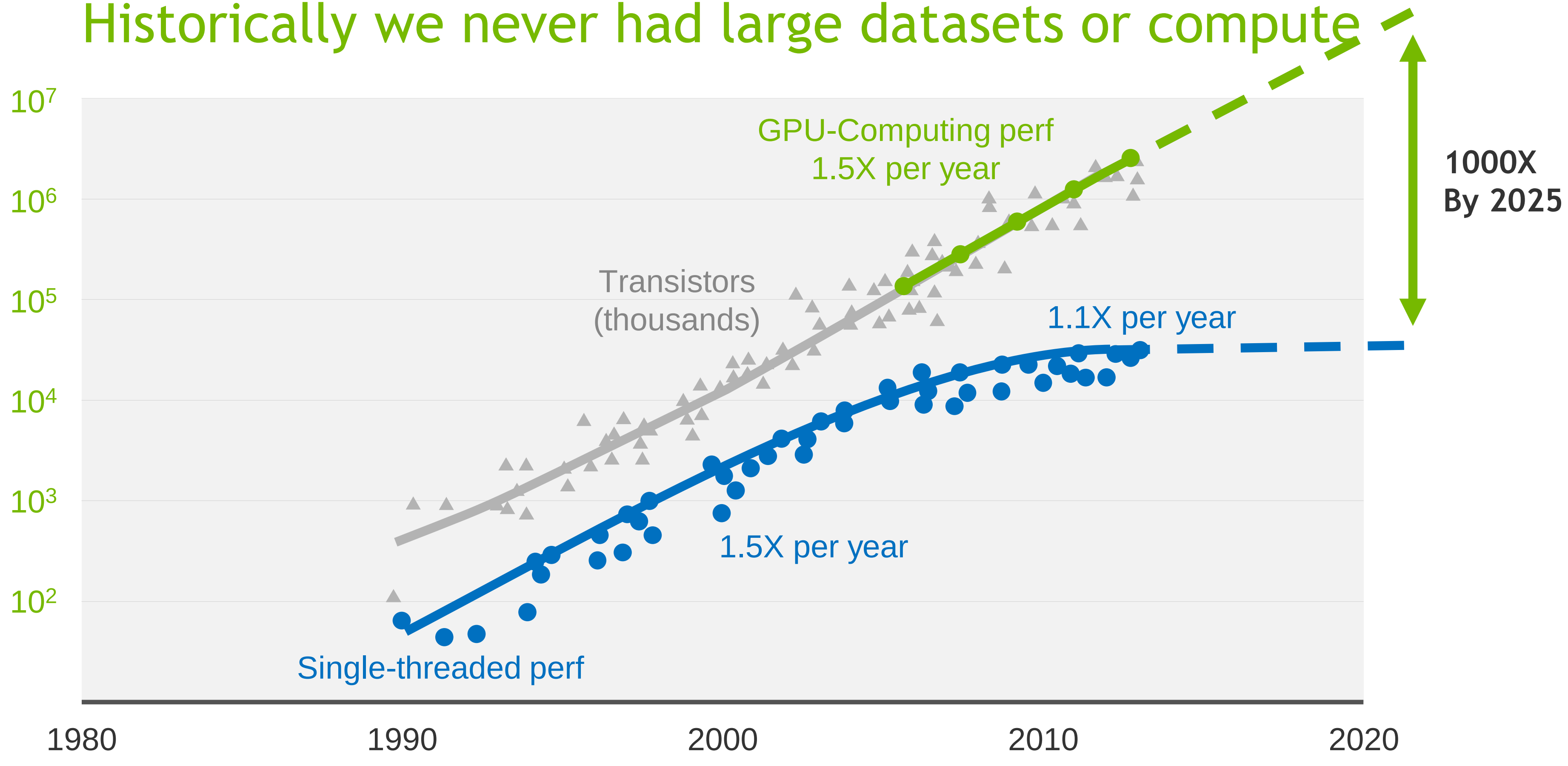




Andrew Ng, “Nuts and Bolts of Applying Deep Learning”, https://www.youtube.com/watch?v=F1ka6a13S9I

COMPUTE

Historically we never had large datasets or compute

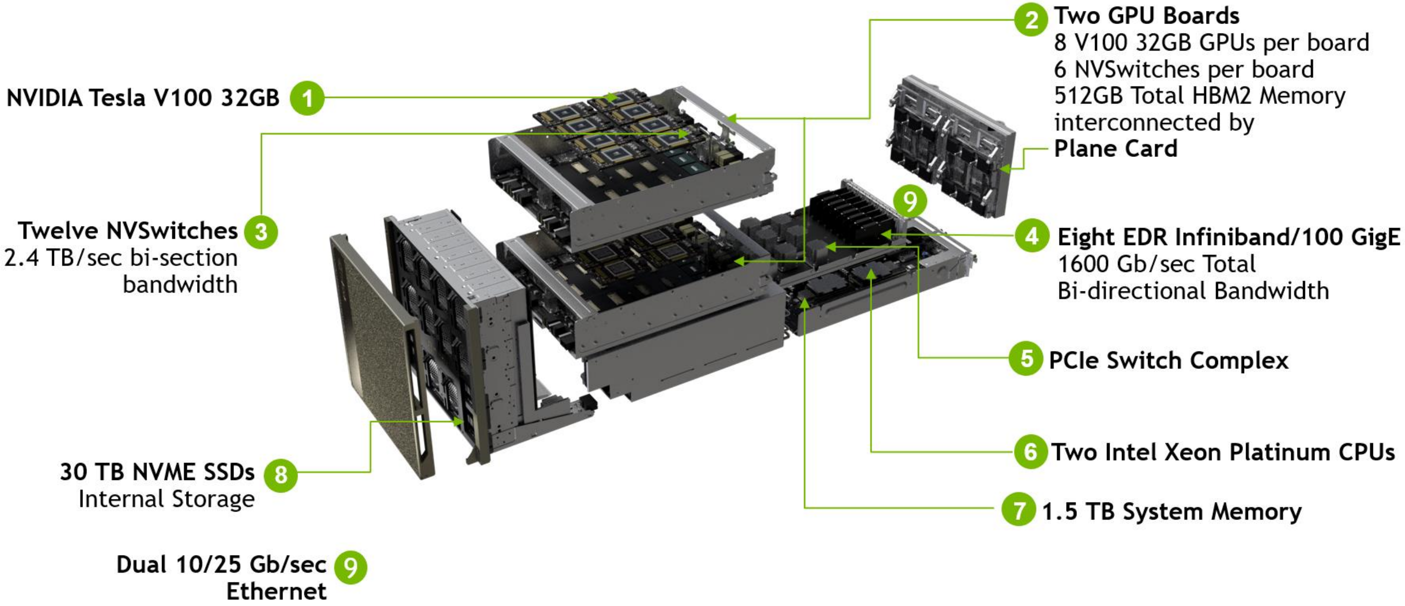


CONTEXT

1.759 petaFLOPs in November 2009

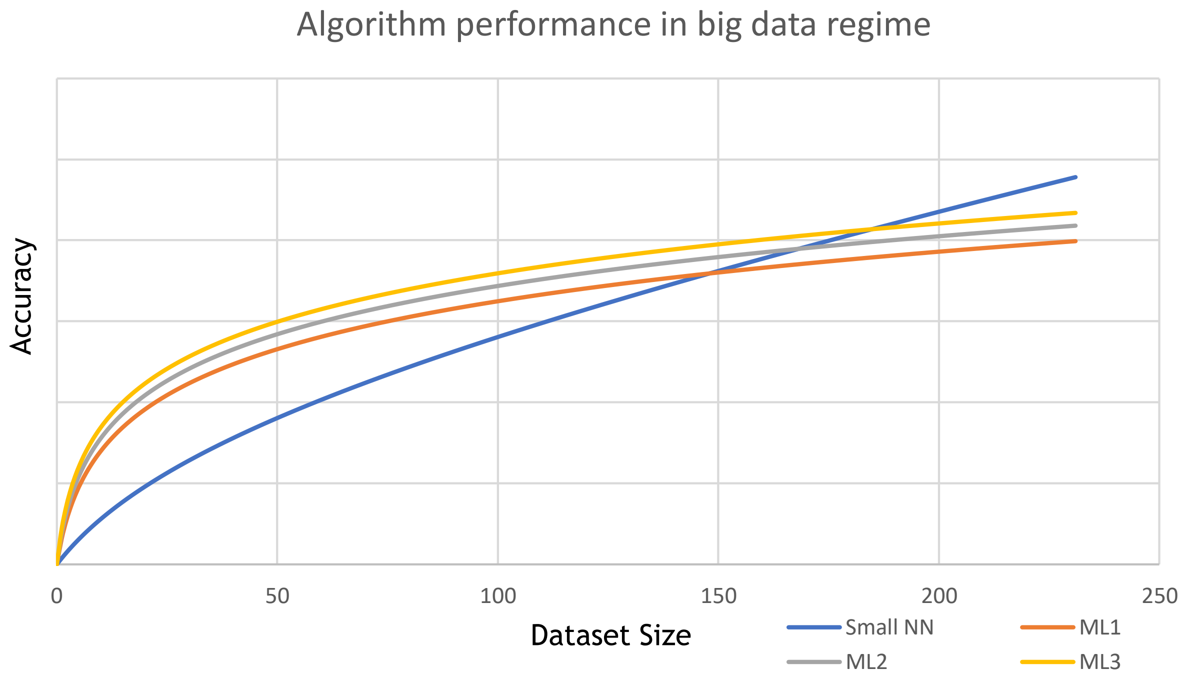
CONTEXT

2 petaFLOPs - today



NEURAL NETWORKS ARE NOT NEW

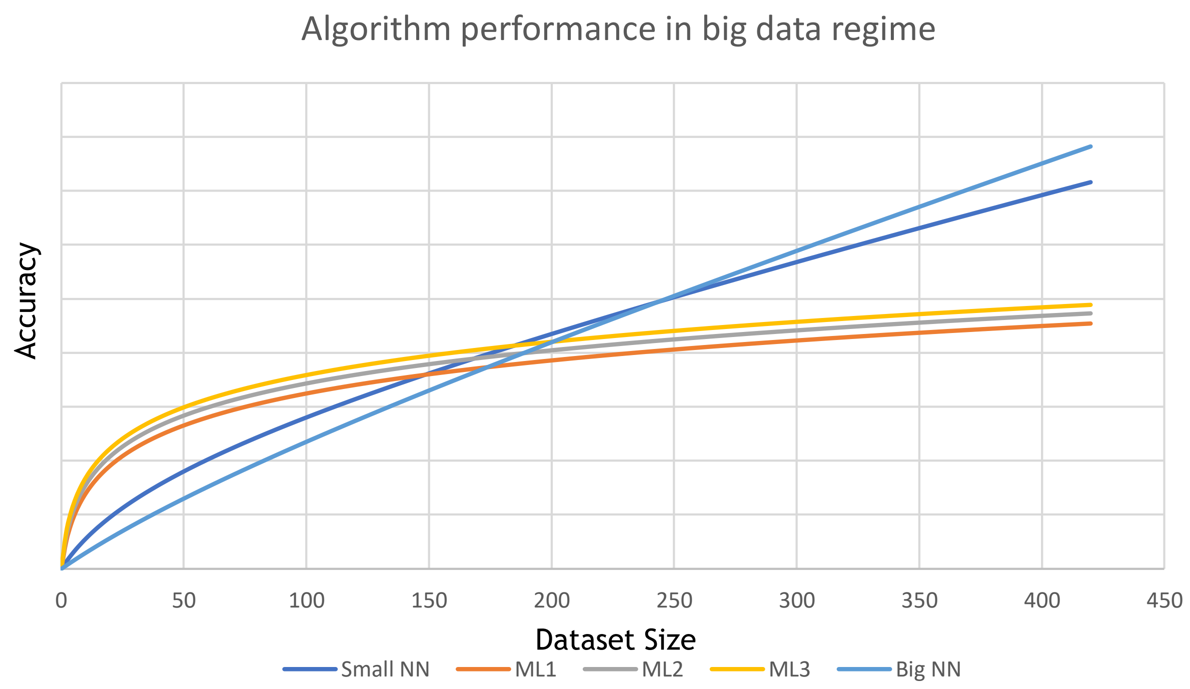
But that changed and transformed the way we do machine learning



Andrew Ng, “Nuts and Bolts of Applying Deep Learning”, https://www.youtube.com/watch?v=F1ka6a13S9I

NEURAL NETWORKS ARE NOT NEW

Data and model size the key to accuracy

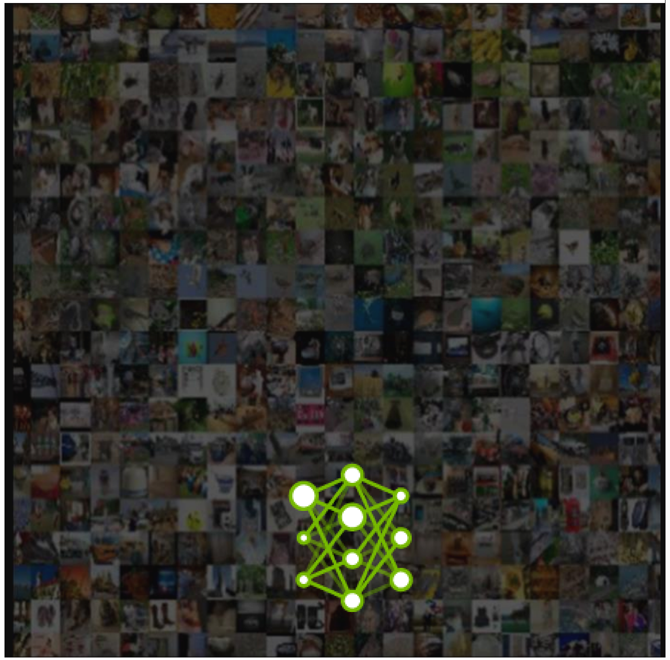


DEEP LEARNING INSTITUTE

NEURAL NETWORK COMPLEXITY IS EXPLODING

To Tackle Increasingly Complex Challenges

7 ExaFLOPS
60 Million Parameters

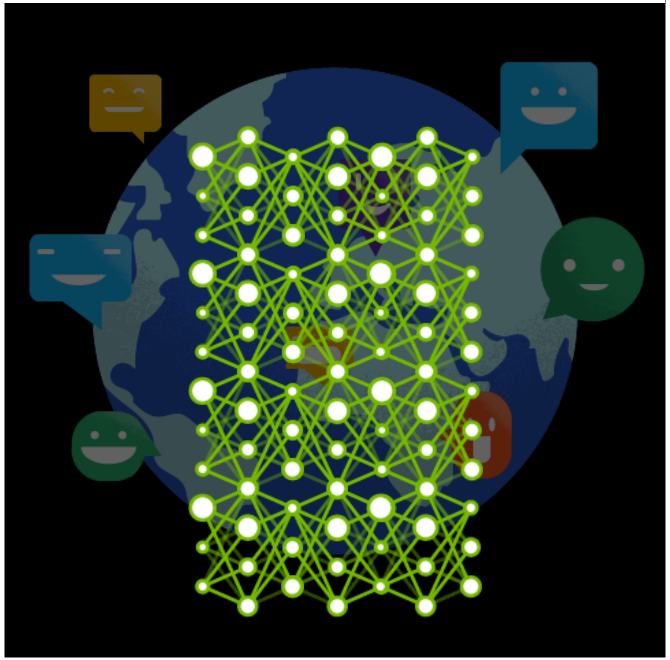


2015 - Microsoft ResNet
Superhuman Image Recognition

20 ExaFLOPS
300 Million Parameters

2016 - Baidu Deep Speech 2
Superhuman Voice Recognition

100 ExaFLOPS
8700 Million Parameters



2017 - Google Neural Machine Translation
Near Human Language Translation

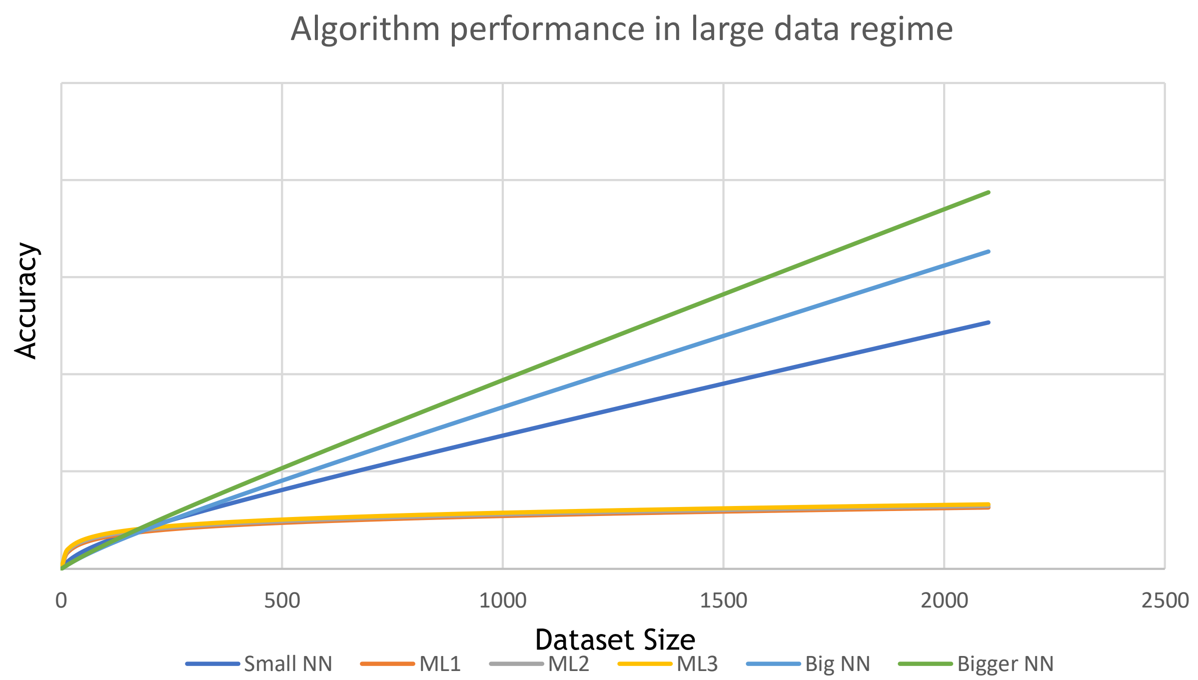
100 EXAFLOPS

=

2 YEARS ON A DUAL CPU SERVER

NEURAL NETWORKS ARE NOT NEW

Exceeding human level performance



Andrew Ng, “Nuts and Bolts of Applying Deep Learning”, https://www.youtube.com/watch?v=F1ka6a13S9I

EXPLODING DATASETS

Logarithmic relationship between the dataset size and accuracy

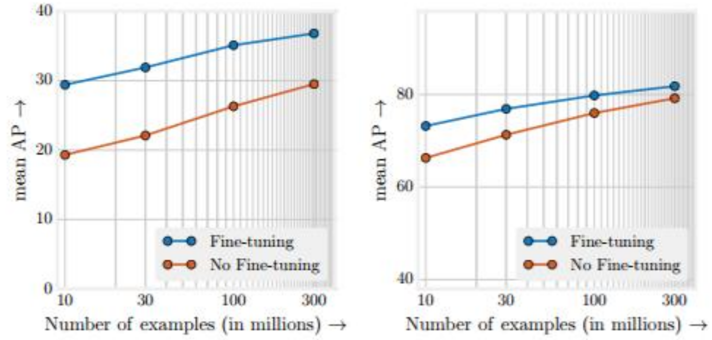


Figure 4. Object detection performance when initial checkpoints are pre-trained on different subsets of JFT-300M from scratch. x-axis is the data size in log-scale, y-axis is the detection performance in mAP@[.5,.95] on COCO minival* (left), and in mAP@.5 on PASCAL VOC 2007 test (right).

Initialization	mIOU
ImageNet	73.6
300M	75.3
ImageNet+300M	**76.5**

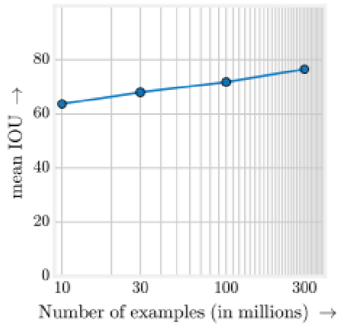
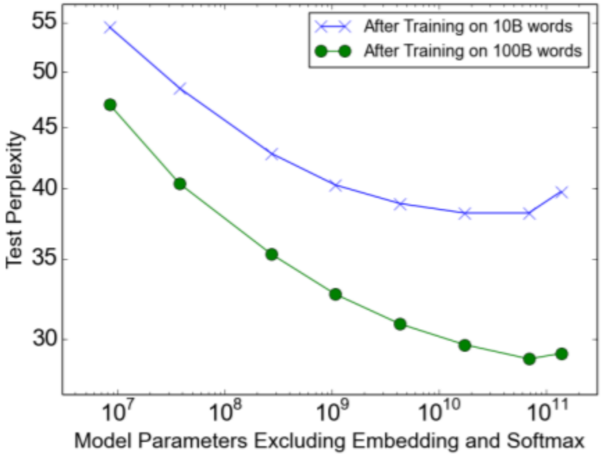


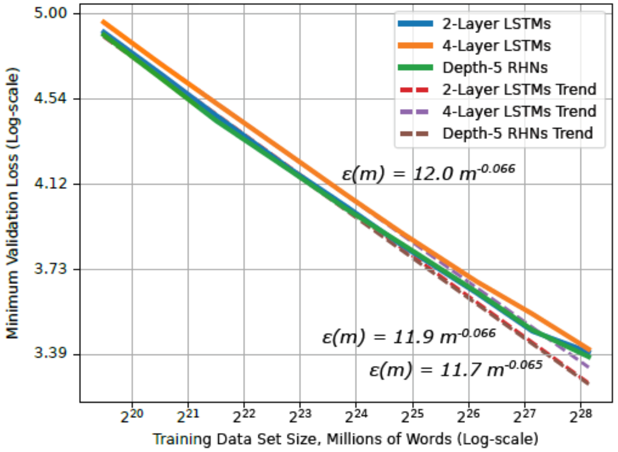
Figure 6. Semantic segmentation performance on Pascal VOC 2012 val set. (left) Quantitative performance of different initializations; (right) Impact of data size on performance.



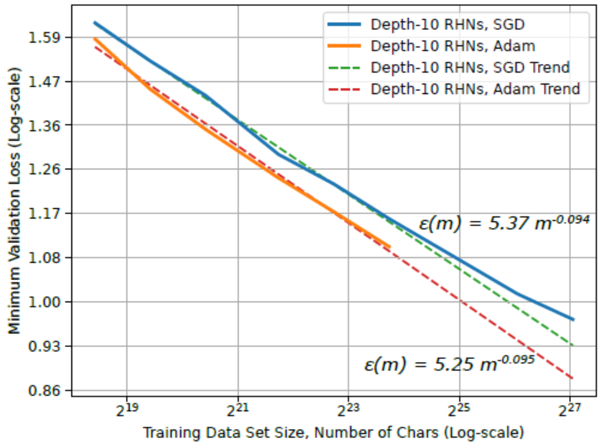
Sun, Chen, et al. "Revisiting Unreasonable Effectiveness of Data in Deep Learning Era." *arXiv preprint arXiv:1707.02968* (2017).
Shazeer, Noam, et al. "Outrageously large neural networks: The sparsely-gated mixture-of-experts layer." arXiv preprint arXiv:1701.06538 (2017).
Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

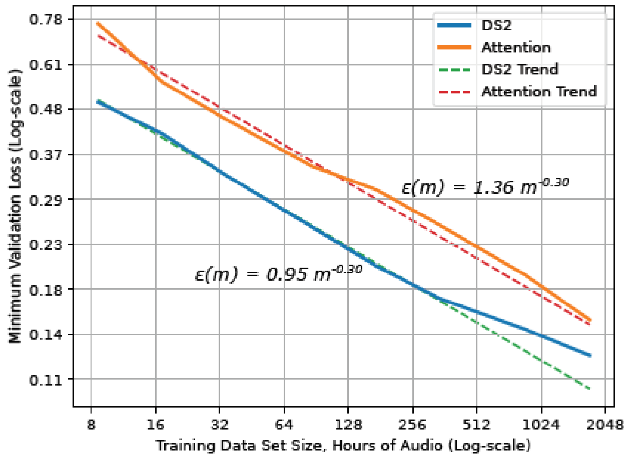
EXPLODING DATASETS

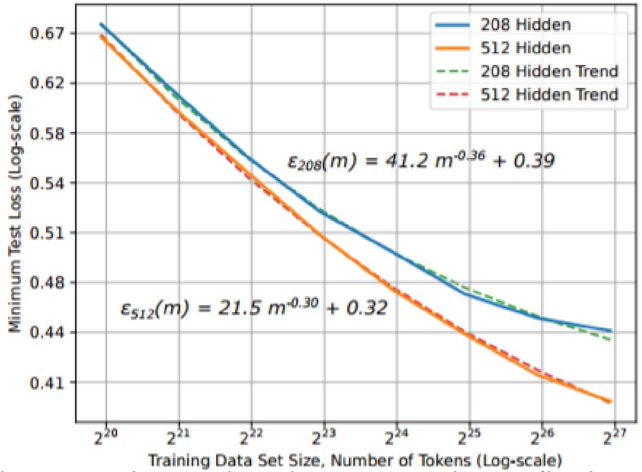
Logarithmic relationship between the dataset size and accuracy

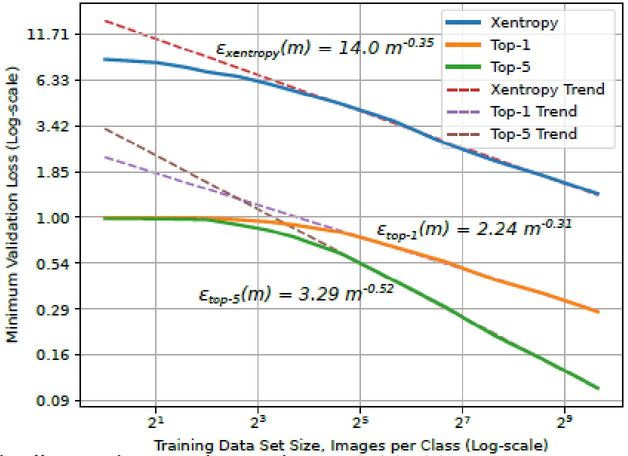


- Translation
- Language Models
- Character Language Models
- Image Classification
- Attention Speech Models





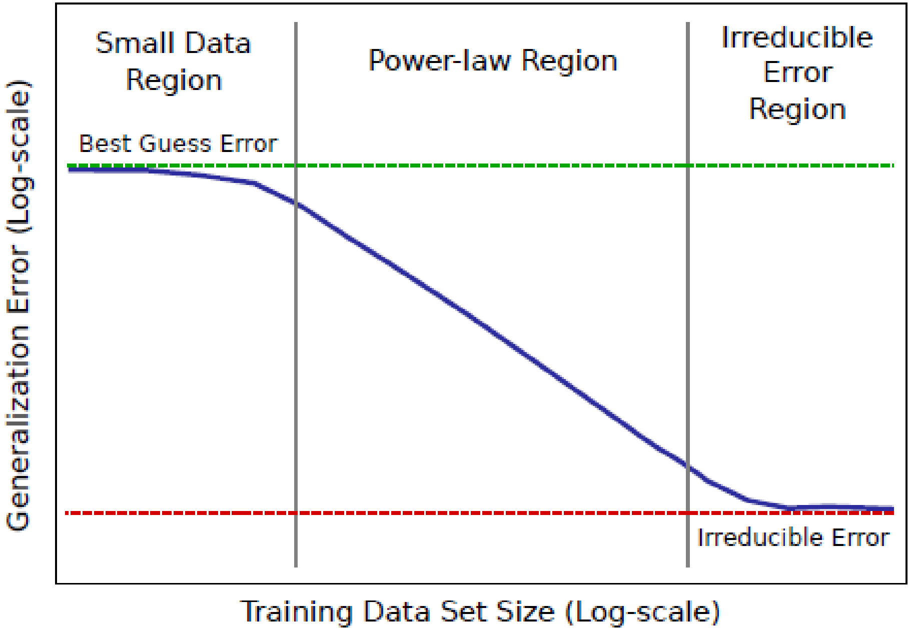




Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., ... & Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. arXiv preprint arXiv:1712.00409.

EXPLODING DATASETS

Logarithmic relationship between the dataset size and accuracy



Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., ... & Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. arXiv preprint arXiv:1712.00409.

Making complex problems easy

Making unsolvable problems expensive

PERCEPTION ALGORITHMS

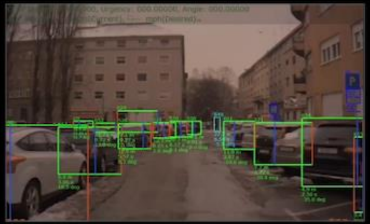
BUILDING AI FOR SDC IS HARD

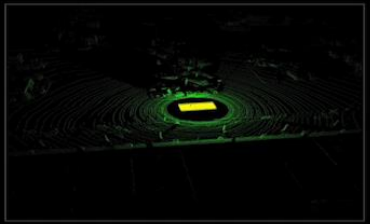
Every neural net in our DRIVE Software stack needs to handle 1000s of conditions and geolocations


Perception

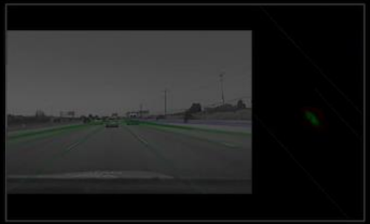

Free Space Perception

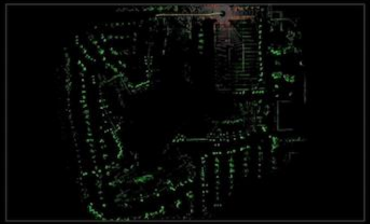

Distance Perception


Weather


LIDAR Perception


Camera-based Mapping


Camera Localization to HD Map


LIDAR Localization to HD Map


Path Perception


Scene Perception

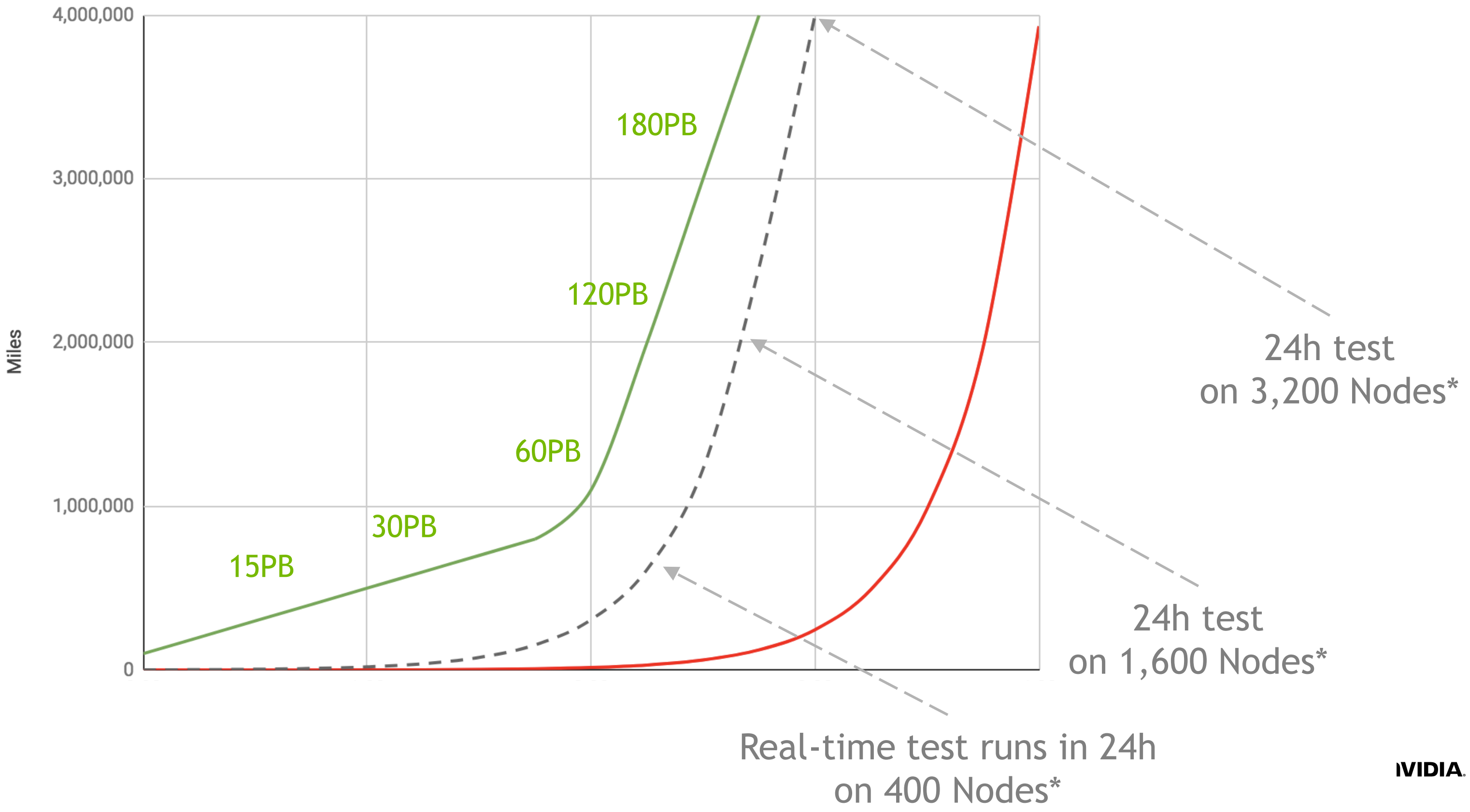
WHAT TESTING SCALE ARE WE TALKING ABOUT?

We're on our way to 100s PB of real test data = millions of real miles
+ 1,000s DRIVE Constellation nodes for offline testing alone
& billions of simulated miles

– Target robustness per model (miles)

-- Test dataset size required (miles)

– NVIDIA's ongoing data collection (miles)



* DRIVE PEGASUS Nodes

NVIDIA

SDC SCALE TODAY AT NVIDIA

12-camera+Radar+Lidar RIG mounted on 30 cars	1,500 labelers	4,000 GPUs in cluster = 500 PFLOPs
1PB collected/week	20M objects labeled/mo	100 DRIVE Pegasus in cluster (Constellations)
15PB active training+test dataset	20 unique models 50 labeling tasks	1PB of in-rack object cache per 72 GPUs, 30PB provisioned

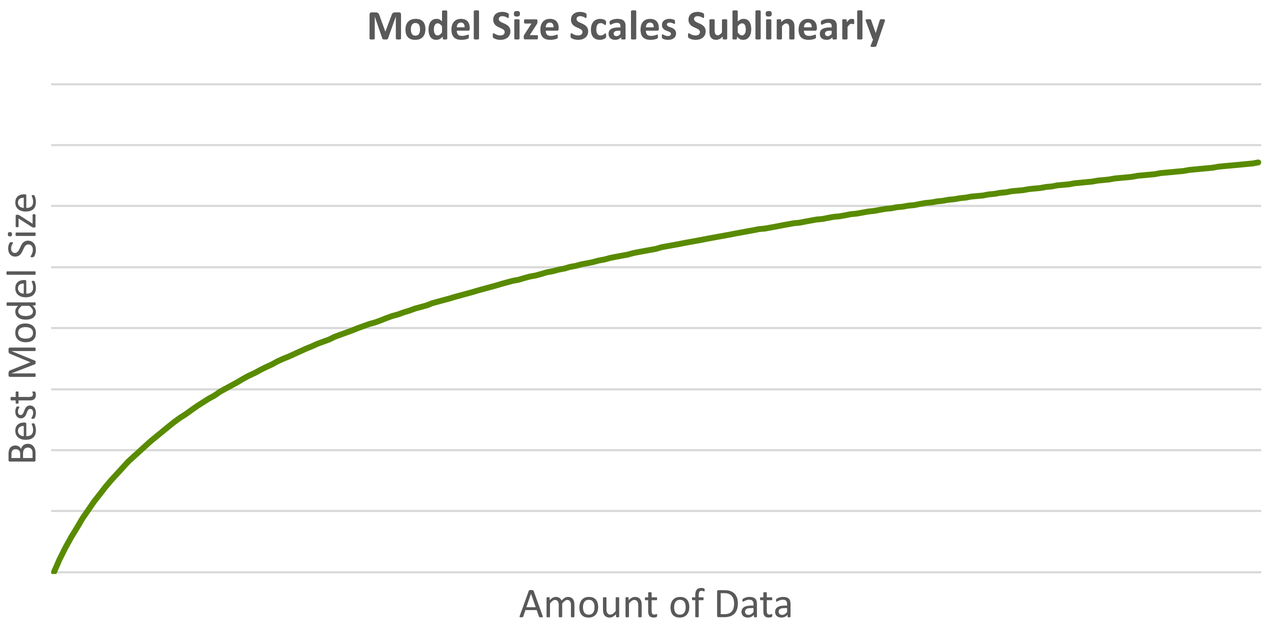
TRAINING

“For any size of the data it’s a good idea to always make the data look small by using a huge model.”

Geoffrey Hinton

EXPLODING MODEL COMPLEXITY

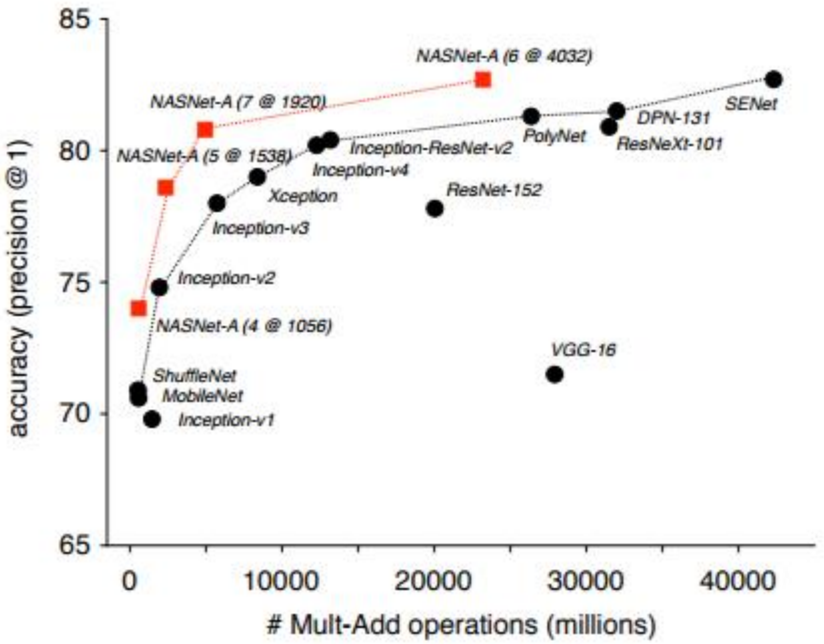
Good news – model size scales sublinearly

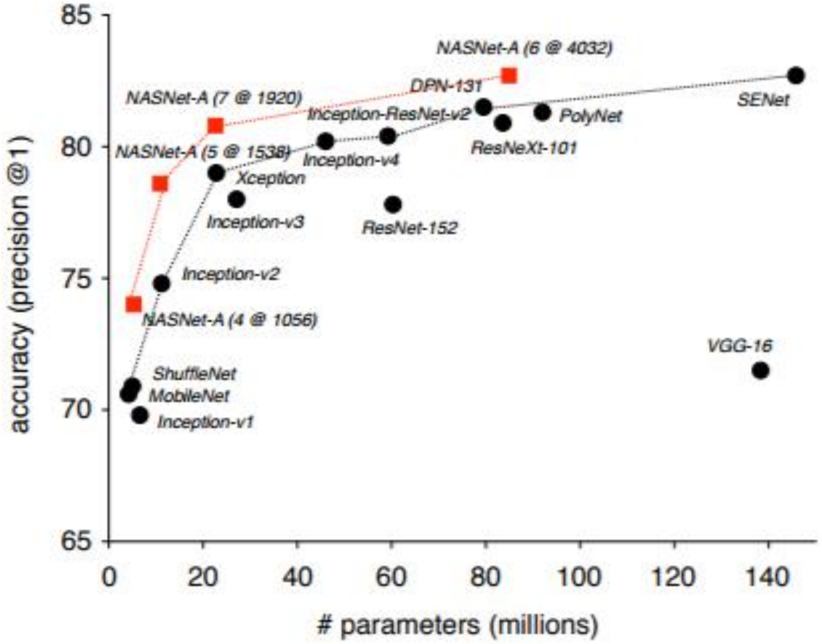


Hestness, J., Narang, S., Ardalani, N., Diamos, G., Jun, H., Kianinejad, H., ... & Zhou, Y. (2017). Deep Learning Scaling is Predictable, Empirically. arXiv preprint arXiv:1712.00409.

EVIDENCE FROM IMAGE PROCESSING

Good news – model size scales sublinearly

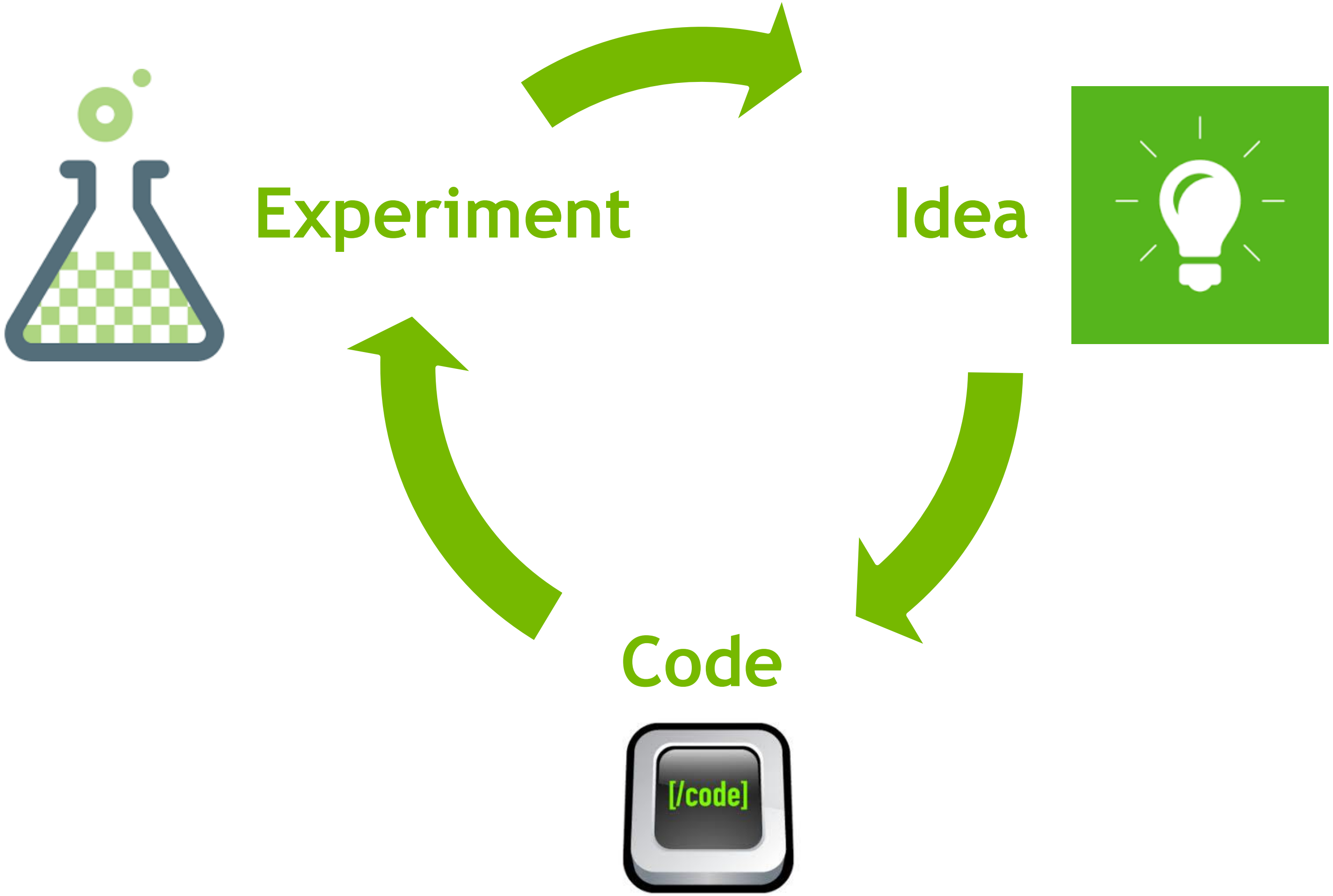




Zoph, Barret, et al. "Learning transferable architectures for scalable image recognition." arXiv preprint arXiv:1707.07012 (2017).

IMPLICATIONS

Experimental Nature of Deep Learning - Unacceptable training time



IMPLICATIONS

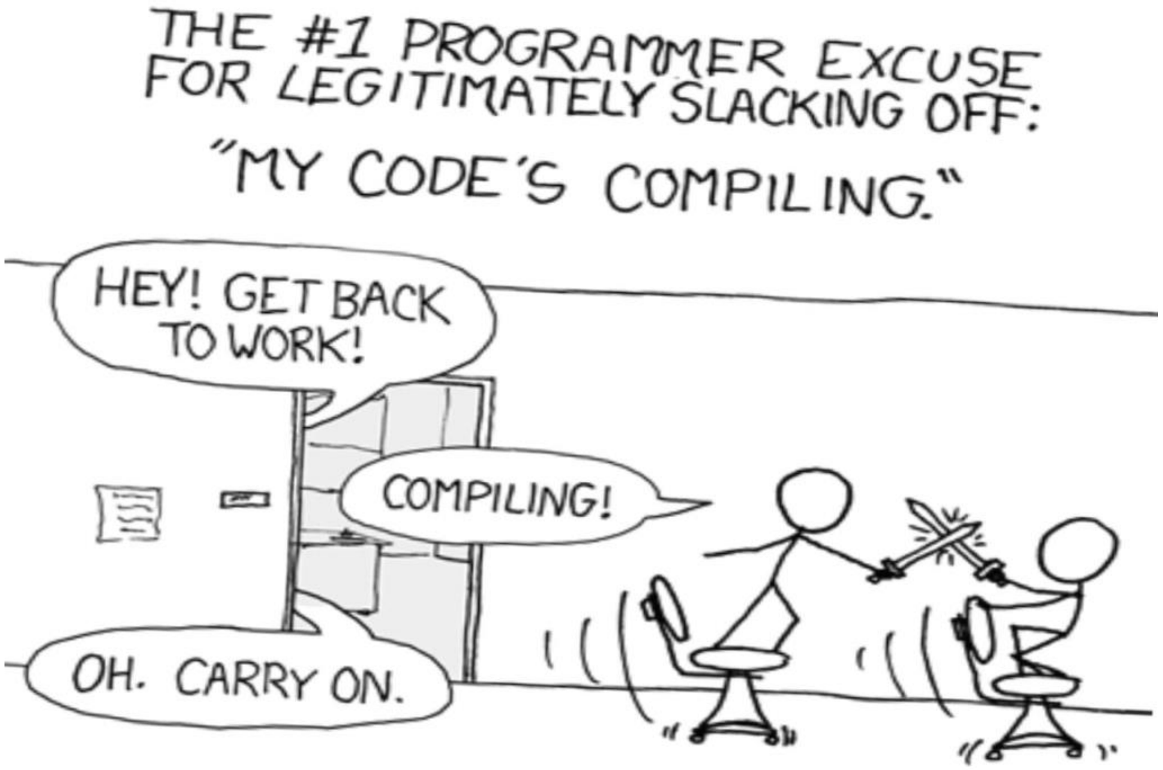
Automotive example

Majority of useful problems are too complex for a single GPU training

	VERY CONSERVATIVE	CONSERVATIVE
Fleet size (data capture per hour)	100 cars / 1TB/hour	125 cars / 1.5TB/hour
Duration of data collection	260 days * 8 hours	325 days * 10 hours
Data Compression factor	0.0005	0.0008
Total training set	104 TB	487.5 TB
InceptionV3 training time (with 1 Pascal GPU)	9.1 years	42.6 years
AlexNet training time (with 1 Pascal GPU)	 1.1 years 	 5.4 years 

CONCLUSIONS

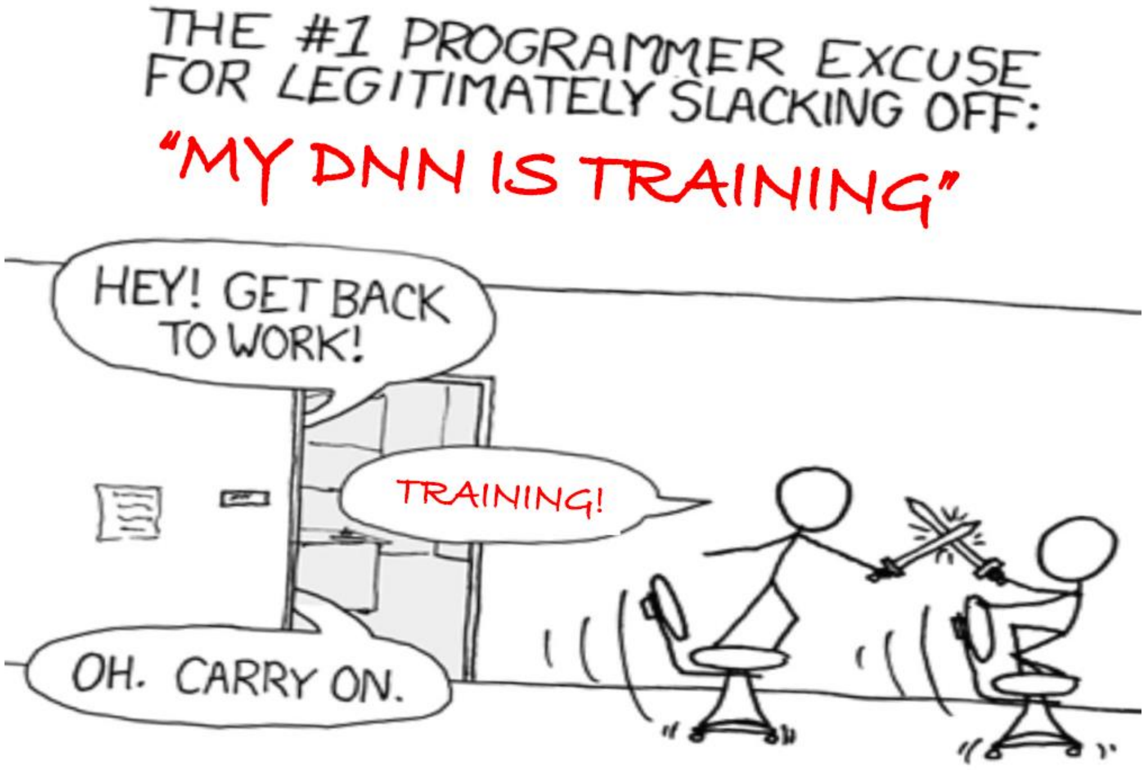
What does your team do in the mean time



Xkcd.com

CONCLUSIONS

What does your team do in the meantime?



Xkcd.com

CONCLUSIONS

Need to scale the training process for a single job

	VERY CONSERVATIVE	CONSERVATIVE
Total training set	104 TB	487.5 TB
InceptionV3 (one DGX-1V)	166 days (5+ **months**)	778 days (2+ **years**)
AlexNet (one DGX-1V)	21 days (3 **weeks**)	98 days (3 **months**)
InceptionV3 (10 DGX-1V's)	16 days (2+ **weeks**)	77 days (11 **weeks**)
AlexNet (10 DGX-1V's)	2.1 days	9.8 days

1 NVIDIA DGX-1

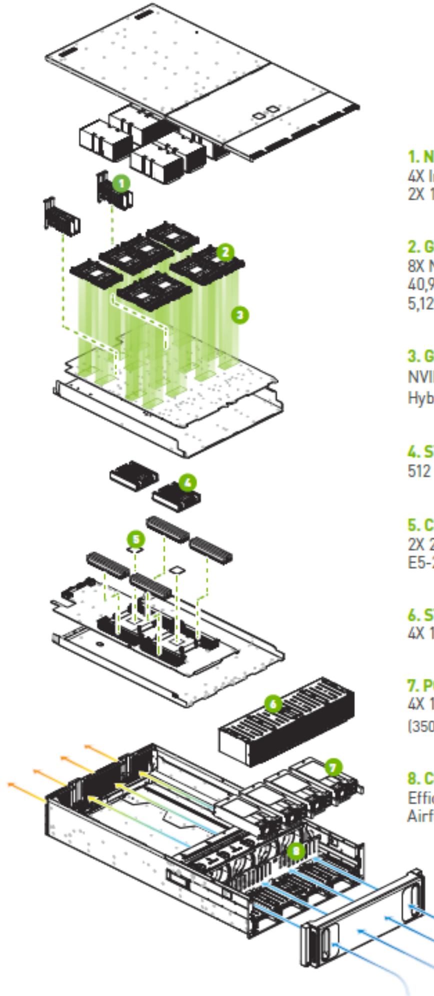
10 NVIDIA DGX-1's

Training
From
Months or Years

To
Weeks or Days

BALANCED HARDWARE

DGX-1 as a reference point for solution design



1. NETWORK INTERCONNECT
4X InfiniBand 100 Gbps EDR
2X 10 GbE

2. GPUs
8X NVIDIA Tesla® V100 16 GB/GPU
40,960 Total NVIDIA CUDA® Cores
5,120 Tensor Cores

3. GPU INTERCONNECT
NVIDIA NVLink™
Hybrid Cube Mesh

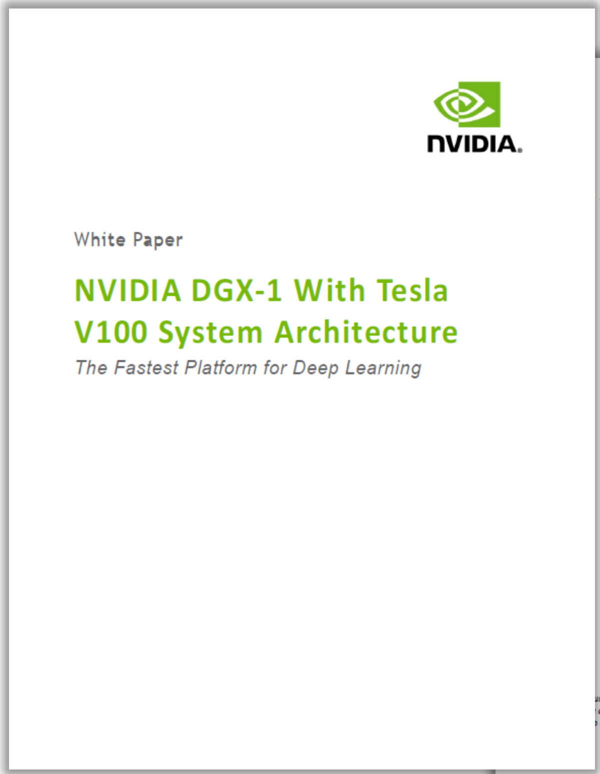
4. SYSTEM MEMORY
512 GB DDR4 LRDIMM

5. CPUs
2X 20-Core Intel® Xeon®
E5-2698 v4 2.2 GHz

6. STREAMING CACHE
4X 1.92 TB SSDs RAID 0

7. POWER
4X 1600 W PSUs
(3500 W TDP)

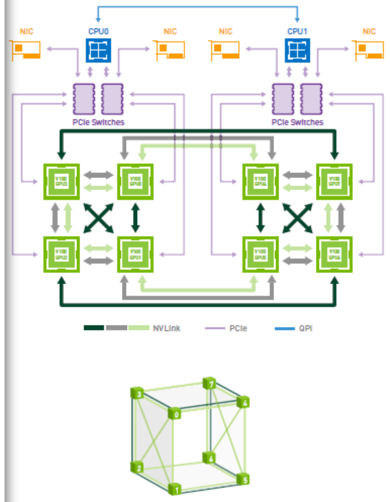
8. COOLING
Efficient Front-to-Back
Airflow



White Paper

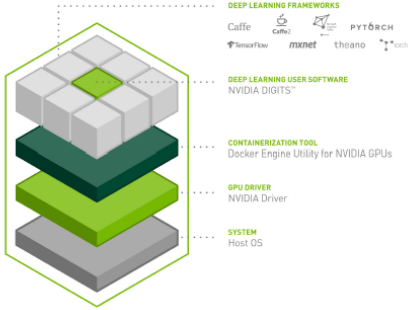
NVIDIA DGX-1 With Tesla V100 System Architecture

The Fastest Platform for Deep Learning



...re 4 DGX-1 uses an 8-GPU hybrid cube-mesh interconnection network topology. ... corners of the mesh-connected faces of the cube are connected to the PCIe tree network, which ... connects to the CPUs and NICs.

NVIDIA DGX-1 With Tesla V100 System Architecture WP-08437-002_v01 | 9



DGX-1 SOFTWARE

...e DGX-1 software has been built to run deep learning at scale. A key goal is to enable practitioners to ...eploy deep learning frameworks and applications on DGX-1 with minimal setup effort. The design of ...e platform software is centered around a minimal OS and driver install on the server, and provisioning ... all application and SDK software in Docker (see Section 4.2) containers through the DGX Container ...egistry[5], maintained by NVIDIA. Containers available for DGX-1 include multiple optimized deep ...arning frameworks, the NVIDIA DIGITS deep learning training application, third-party accelerated ...lutions, and the NVIDIA CUDA Toolkit. Figure 9 shows the DGX-1 deep learning software stack.

Figure 9 The DGX-1 Deep Learning Software Stack.

5. NVIDIA's Docker container registry service. See http://docs.nvidia.com/dgx/dgx-registry-guide/

...NVIDIA DGX-1 With Tesla V100 System Architecture WP-08437-002_v01 | 18

DGX POD

Reference architectures

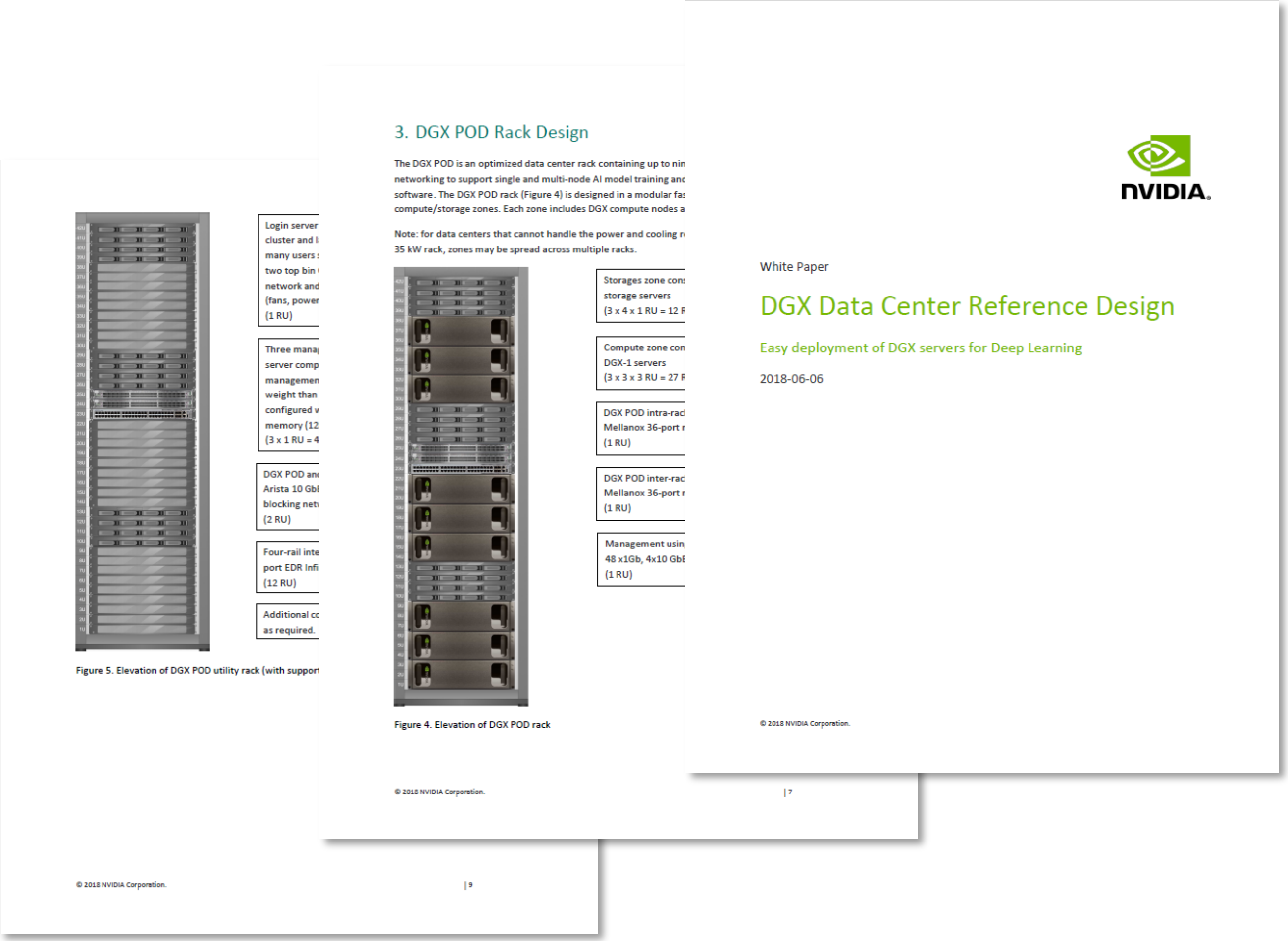

3. DGX POD Rack Design

Figure 4. Elevation of DGX POD rack

Figure 5. Elevation of DGX POD utility rack (with support

White Paper

DGX Data Center Reference Design

Easy deployment of DGX servers for Deep Learning

2018-06-06

© 2018 NVIDIA Corporation.

- *Hiding the complexity of hardware*
- *Hiding complexity of devops*
- *Hiding the complexity of software toolkit management*
- *Partner reference architectures with every major storage provider*

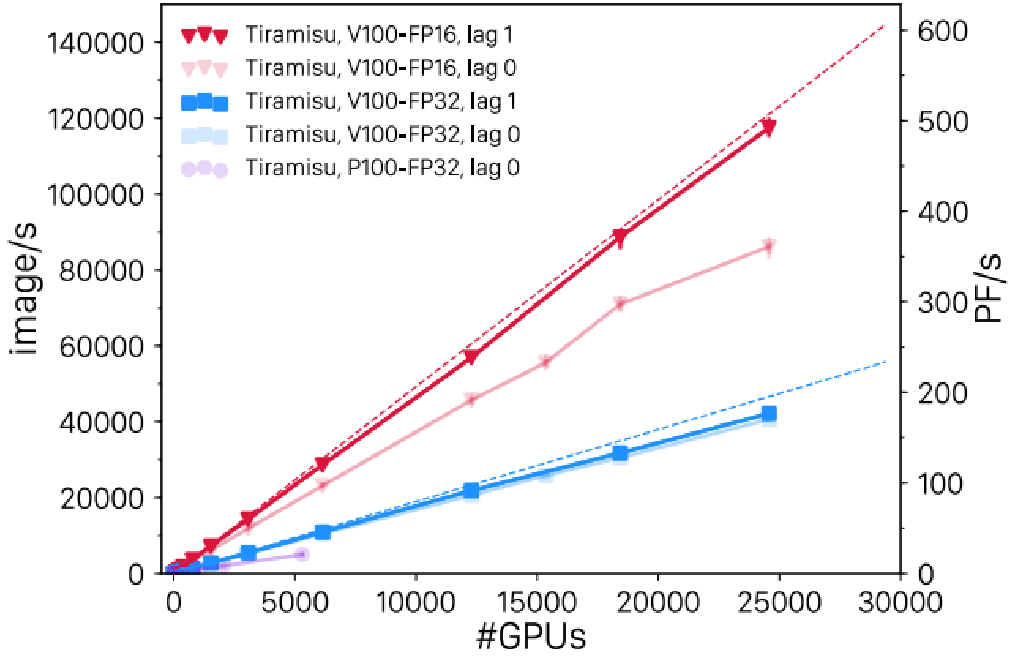
DEEP LEARNING INSTITUTE

TRAINING

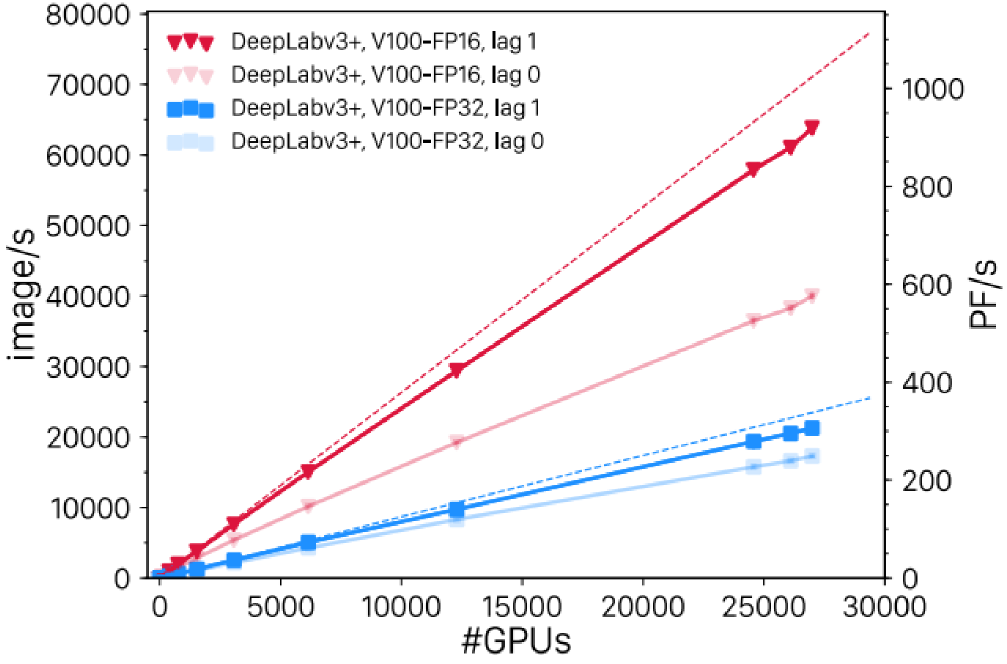
660
NODES
5,280
V100 GPUs
660
petaFLOPS (AI)
1 GPU
80
petaFLOPS (FP32)
40
petaFLOPS (FP64)
15
gigaFLOPS per watt

SUMMIT

27360 Tesla V100



(a) Tiramisu

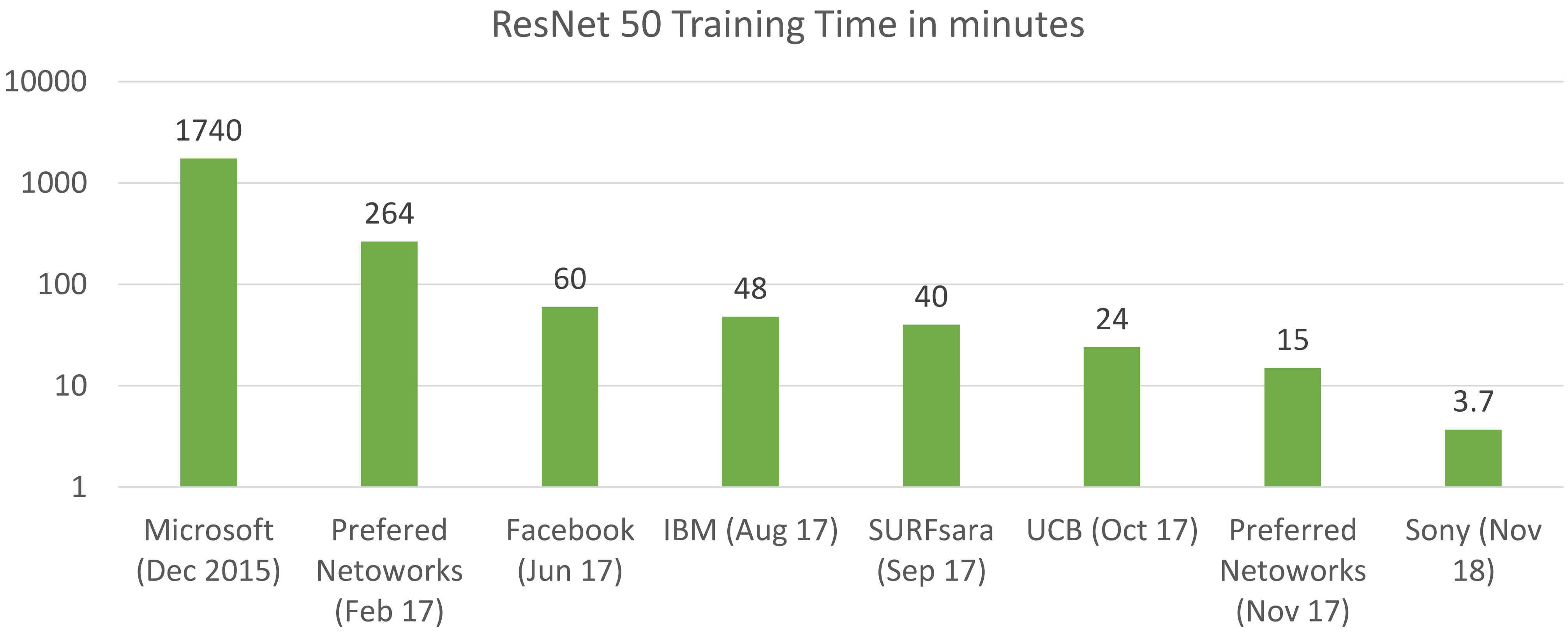


(b) DeepLabv3+

Kurth, T., Treichler, S., Romero, J., Mudigonda, M., Luehr, N., Phillips, E., ... & Houston, M. (2018, November). Exascale deep learning for climate analytics. In Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis (p. 51). IEEE Press.

ITERATION TIME

Short iteration time is fundamental for success



VALIDATION

SAFE AV REQUIRES A COMPREHENSIVE VALIDATION APPROACH

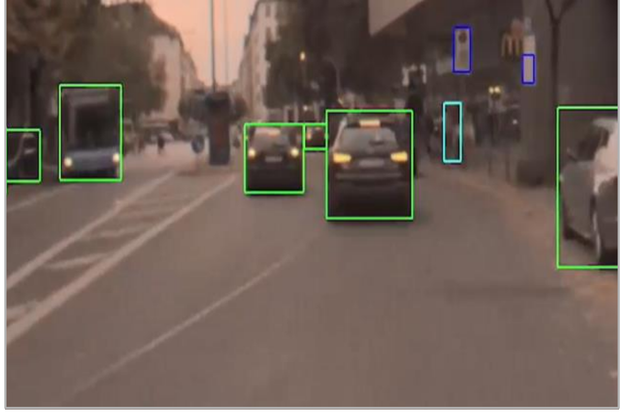
Large Scale | Millions of Miles

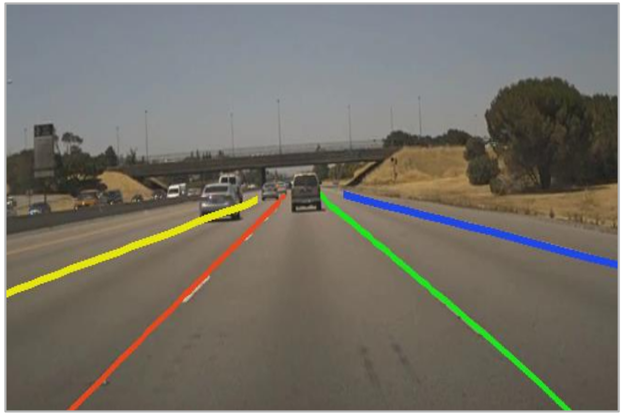
Diverse Vehicle and World Conditions

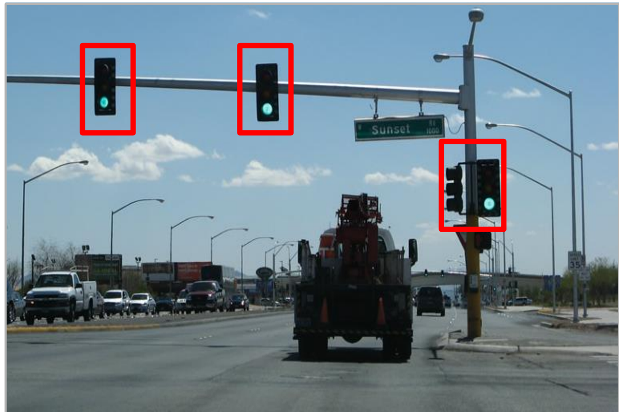
Data Driven | Scenario based

Repeatable and Reproducible

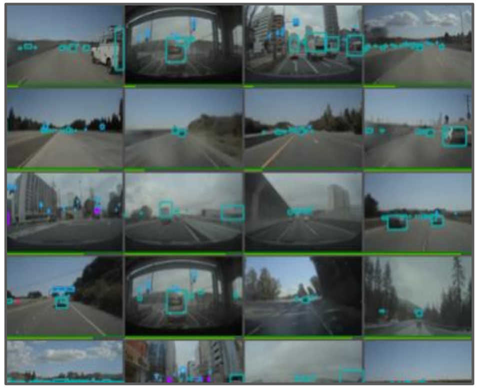
End-to-End System Level Test







THE AV VALIDATION GAP



COMPONENT LEVEL SIL

Low Fidelity | Scalable

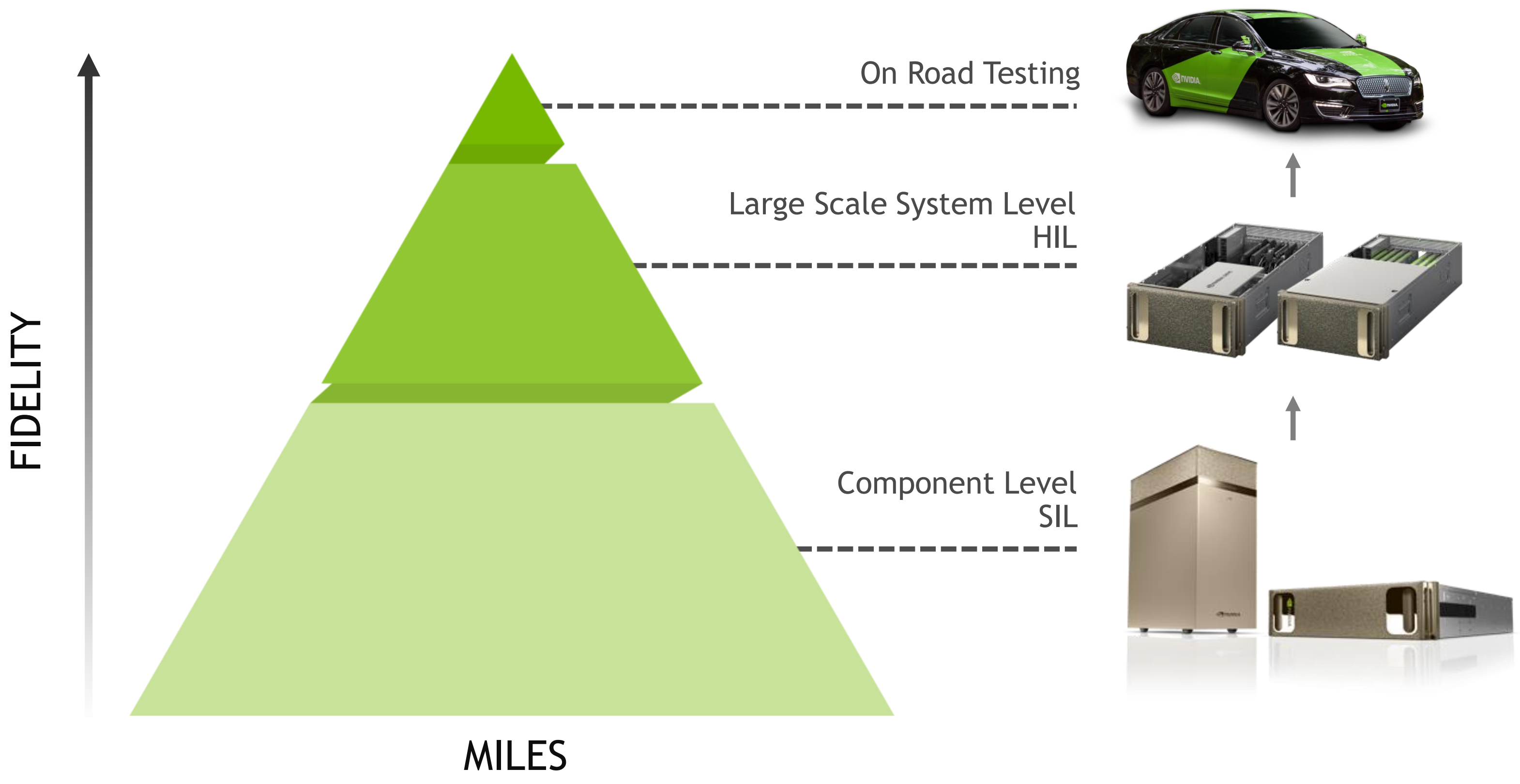
ON ROAD TESTING

High Fidelity| Doesn't Scale

No Coverage for Extreme & Dangerous Scenarios

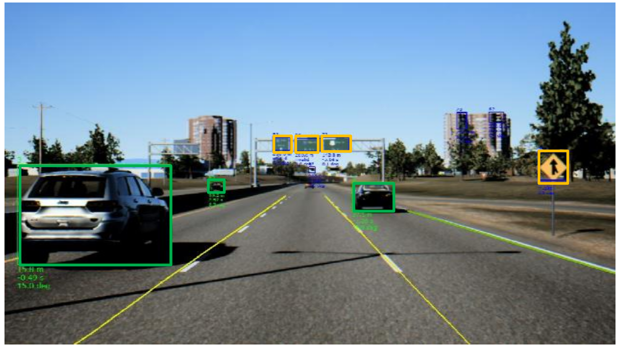
NVIDIA DRIVE VALIDATION METHODOLOGY

Three Pronged Approach



HARDWARE IN THE LOOP SIMULATION

Bit Accurate & Timing Accurate



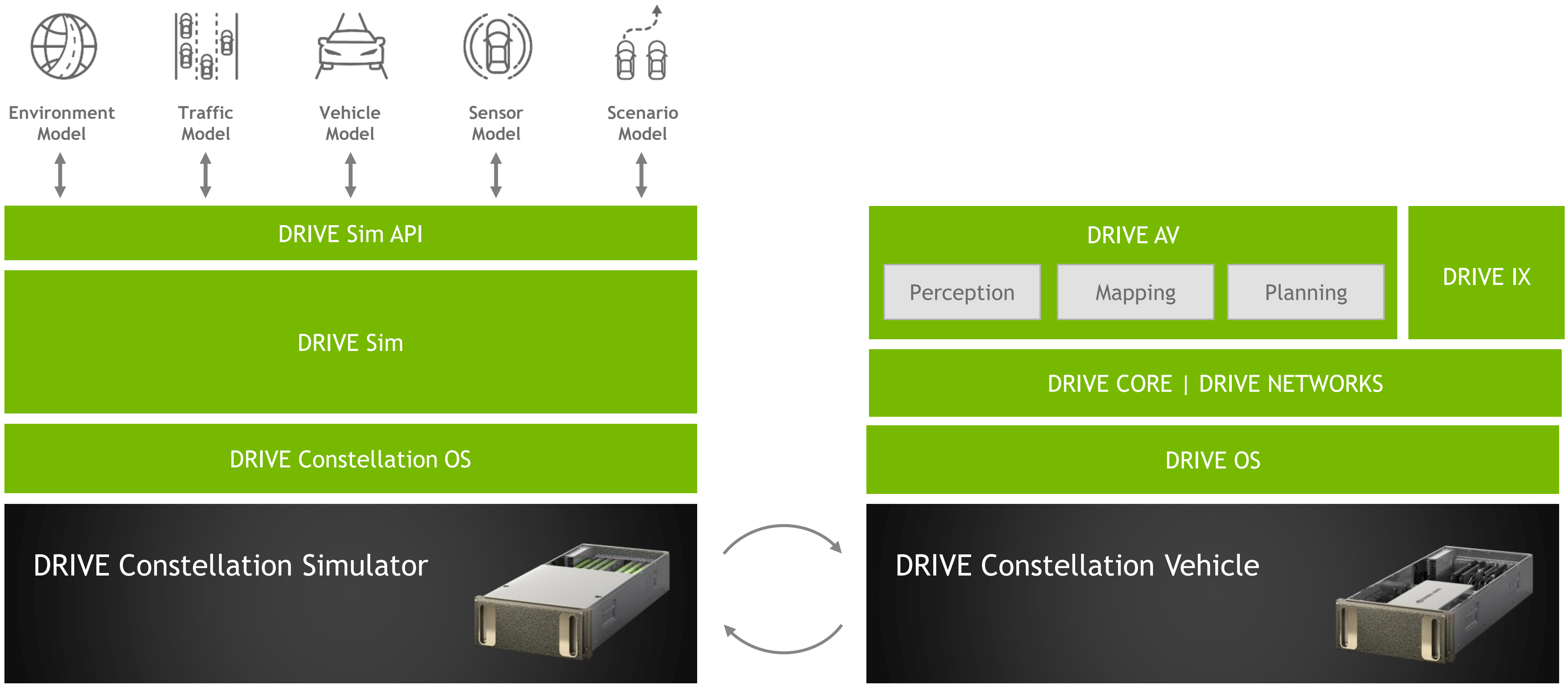
PERCEPTION

Camera | Radar | Lidar | IMU

CONTROL

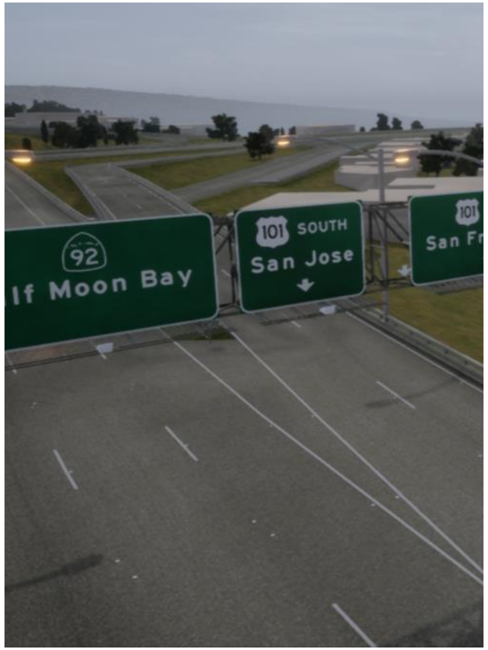
Steering | Throttle | Brake

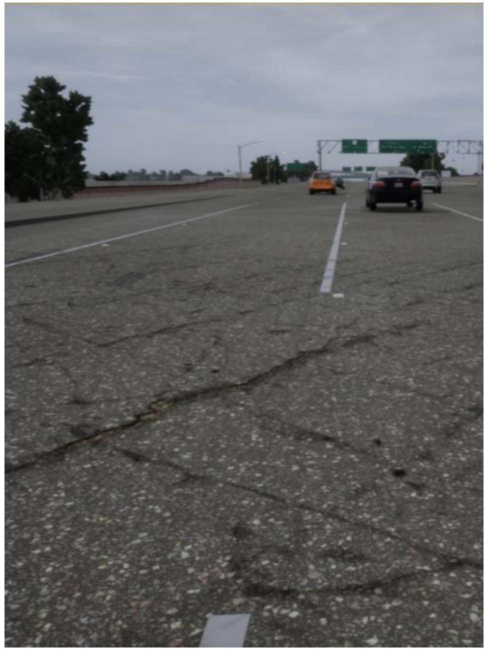
DRIVE CONSTELLATION ARCHITECTURE



ENVIRONMENT MODEL

Highly Detailed Environments




ENVIRONMENT MODEL

Change Time of Day

ENVIRONMENT MODEL

Add Traffic Scenarios

ENVIRONMENT MODEL

Change Weather

DRIVE CONSTELLATION FOR 3RD PARTY AV VALIDATION

Open | Accessible | Available at Scale

Open Platform | Wide Ecosystem Support

Cloud Based Solution | Scalable

Accessible to OEMs and Researchers

Demonstrate Best Practice for AV Validation

THE DRIVE SIM ECOSYSTEM

MAPPING

3Dmapping cloud
here
OpenDRIVE
TomTom
ZENRIN

VEHICLE DYNAMICS

carSIM
IPG AUTOMOTIVE
VI-GRADE

ENVIRONMENT MODEL

51VR
parallel domain

SYSTEM INTEGRATORS

AVL
dSPACE

SENSOR MODEL

ANSYS
ECHODYNE
esi
INNOVIZ TECHNOLOGIES
SONY
SIEMENS
Velodyne

SCENARIO MODEL

AAi
cognata
rFpro
VIRES

TRAFFIC MODEL

cognata
PTV VISUM
SUMO

3rd PARTY CERTIFICATION

CATARC
IIHS HLDI
TÜV SÜD

BEST PRACTICES FOR AV VALIDATION

Planning the Path to Safety

Partnerships with leading safety organizations
Public | Private | Worldwide
Creating best practices and standards

THANK YOU

twitter.com/nvidiaAI