A Holistic View of Perception in Intelligent Vehicles

Ghassan AlRegib, PhD
Professor

Mohit Prabhushankar, PhD
Postdoctoral Fellow

Omni Lab for Intelligent Visual Engineering and Science (OLIVES)
School of Electrical and Computer Engineering
Georgia Institute of Technology
{ alregib, mohit.p } @gatech.edu

June 04, 2023 – Anchorage, Alaska

IEEE IV 2023
IEEE INTELLIGENT VEHICLES SYMPOSIUM

OLIVES @GeorgiaTech

Georgia Tech

Autonomous Vehicles

Why Autonomous Vehicles?

Safety in Mobility

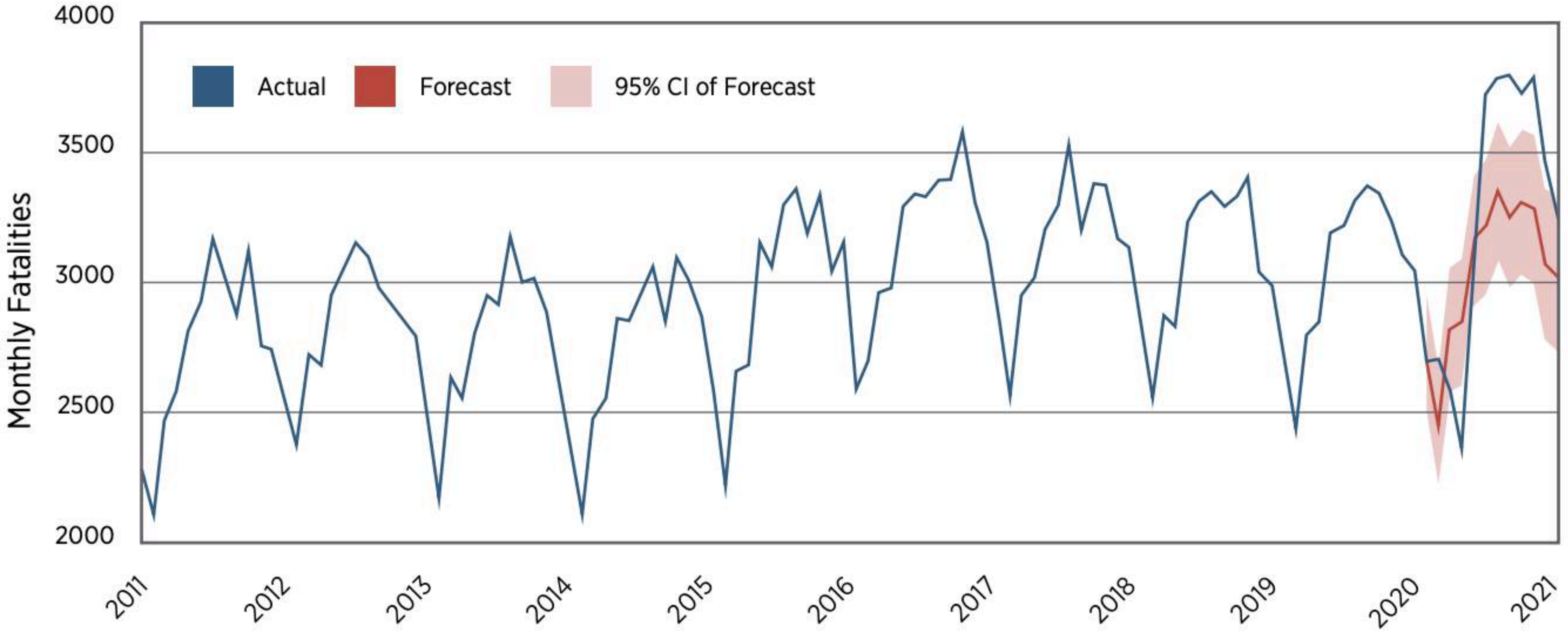
Mobility Experience

Safety: https://www.lensculture.com/articles/arnold-odermatt-karambolage-smash-up#slideshow
Experience: https://innovationatwork.ieee.org/autonomous-vehicles-for-today-and-for-the-future/

Autonomous Vehicles

Why Autonomous Vehicles?

In 2020, despite COVID-19 restrictions, fatalities increased in the US



Tefft, B.C. & Wang, M. (2022). *Traffic Safety Impact of the COVID-19 Pandemic: Fatal Crashes Relative to Pre-Pandemic Trends, United States, May–December 2020* (Research Brief). Washington, D.C.: AAA Foundation for Traffic Safety.

Next Revolution in Mobility Safety: AI

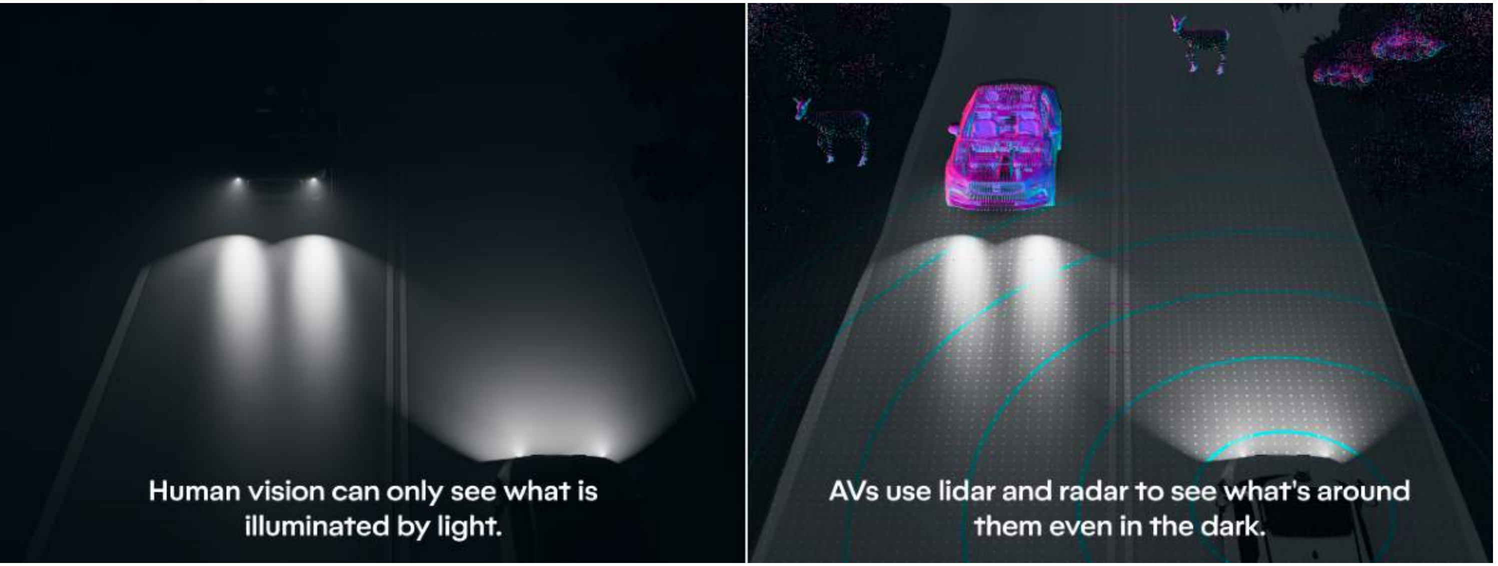
94% of all car accidents are due to human error

It is estimated that, globally, AVs can prevent 4.22 million accidents per year by 2050

Center for Sustainable Systems, University of Michigan. 2021. "Autonomous Vehicles Factsheet." Pub. No. CSS16-18.

Autonomous Vehicles
How will AI ensure Safety in Mobility?

AI identifies and overcomes human limitations in sensing and simulates complex environments for testing



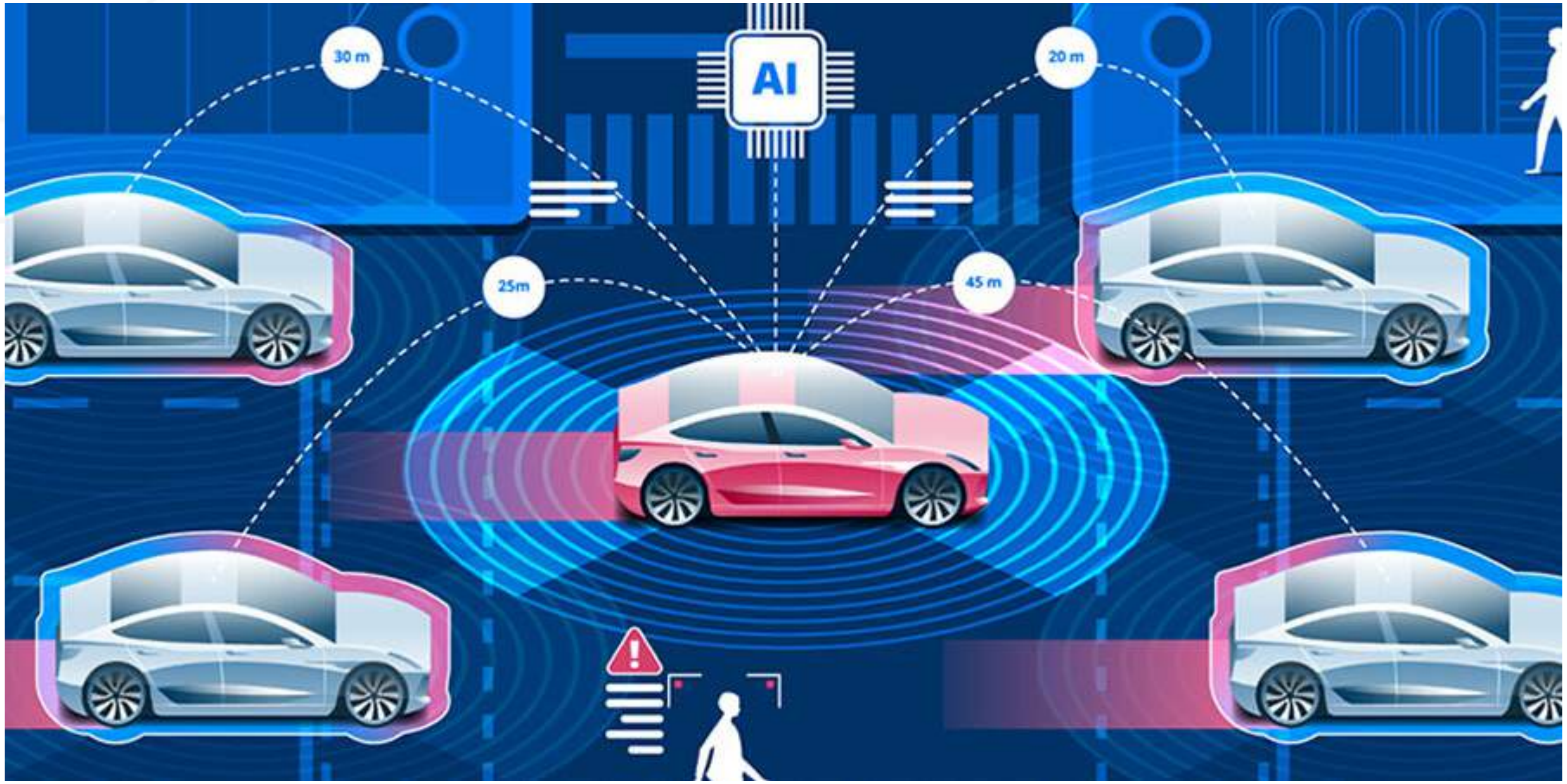
Active sensors like LIDAR overcome the limitations of passive vision sensing

Incredibly complex driving scenarios can be simulated using AI to test itself

https://www.unrealengine.com/en-US/spotlights/multi-purpose-car-simulation-environment-gets-a-boost-from-unreal-engine

Autonomous Vehicles
How will AI ensure Safety in Mobility?

AI provides technologies to handle large data modalities in real time environments



Real-time connection to other vehicles, pedestrians, infrastructure and networks is facilitated by AI

https://www.telecomreview.com/articles/reports-and-coverage/3985-connected-and-autonomous-cars-balancing-morality-and-regulation

Objectives

Objectives of the Tutorial

- Part I: Challenges in Perception and Autonomy
- Part II: Deep Learning for Perception
- Part III: Existing Deep Learning solutions to Challenges in Perception
- Part IV: Remaining Challenges and Future Directions

A Holistic View of Perception in Intel. Vehicles

Part I: Perception and Autonomy

IEEE IV 2023
IEEE INTELLIGENT VEHICLES SYMPOSIUM

OLIVES @GeorgiaTech

Georgia Tech

Objectives

Objectives in Part I

- Summarize the progress of AVs over the years
- Discuss the role of perception in AVs and where it fits within the AV workflow
- Review well-known failures of AVs in providing safety to drivers and to others
- Discuss major technical challenges currently facing AV
- Motivate deep learning as a holistic solution to perception challenges

Perception
What is Perception?

ME!!!!

What is perception?

See, process, understand.

https://www.animalcognition.org/2015/04/15/list-of-animals-that-have-passed-the-mirror-test/

Perception
Perception in AVs

Perception in AVs

Tsubaka Mechanical Engineering Laboratory (1977)

First standalone "autonomous" vehicle


Automatically Operated Car

Technology demonstrated:

Two video cameras and an analog computer onboard for image processing, Detect street markings

Tsubaka: Srinivas Rao, P., Gudla, R., Telidevulapalli, V. S., Kota, J. S., & Mandha, G. (2022). Review on self-driving cars using neural network architectures.

Perception in AVs
Eureka PROMETHEUS Project (1987 - 1995)

New technologies demonstrated:

Vision enhancement, Lane keeping support, visibility range monitoring, Driver status monitoring, Collision avoidance, Cooperative driving, Autonomous intelligent cruise control

PROMETHEUS: https://en.wikipedia.org/wiki/Eureka_Prometheus_Project

Perception in AVs
DARPA Grand Challenge (2004 - 2005)

New technologies demonstrated:

Wide sensor suite including stereo vision, LIDAR, radar, and ultrasound sensors, sensor fusion, obstacle detection, off-road path following, path finding

Urmson, Chris, Charlie Ragusa, David Ray, Joshua Anhalt, Daniel Bartz, Tugrul Galatali, Alexander Gutierrez et al. "A robust approach to high-speed navigation for unrehearsed desert terrain." *Journal of Field Robotics* 23, no. 8 (2006): 467-508.

Georgia Tech in DARPA Challenge
Need for Failsafe in AVs

Video/News Articles

Remote Repositioning

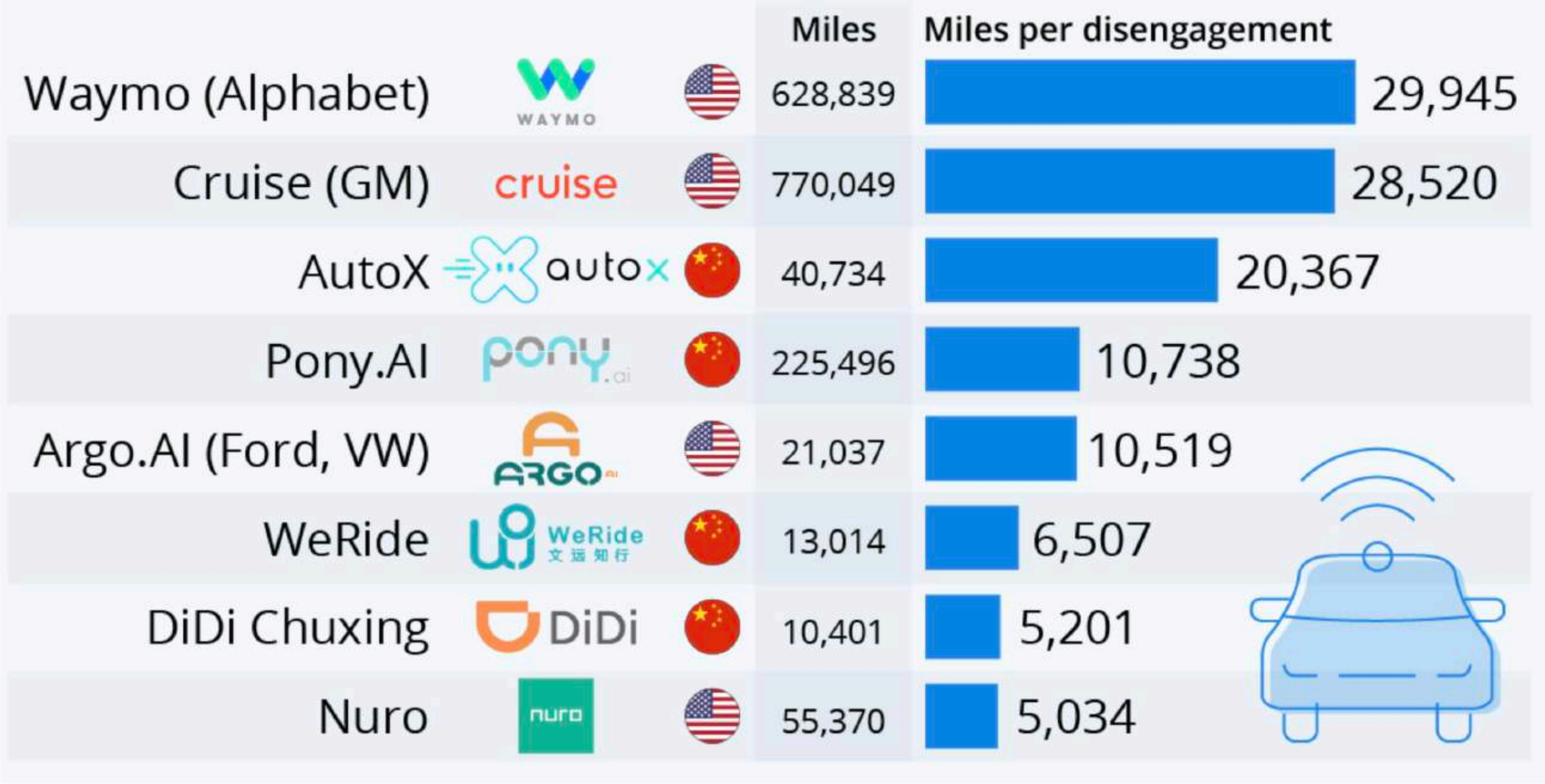
A driver in the Cloud Remotely Drives a Completely Equipped Vehicle

New technologies demonstrated:

Low latency failsafe mechanisms in connected cars

Perception in AVs
A Leap in Progress

AV statistics in California (Dec 2019 – Nov 2020)

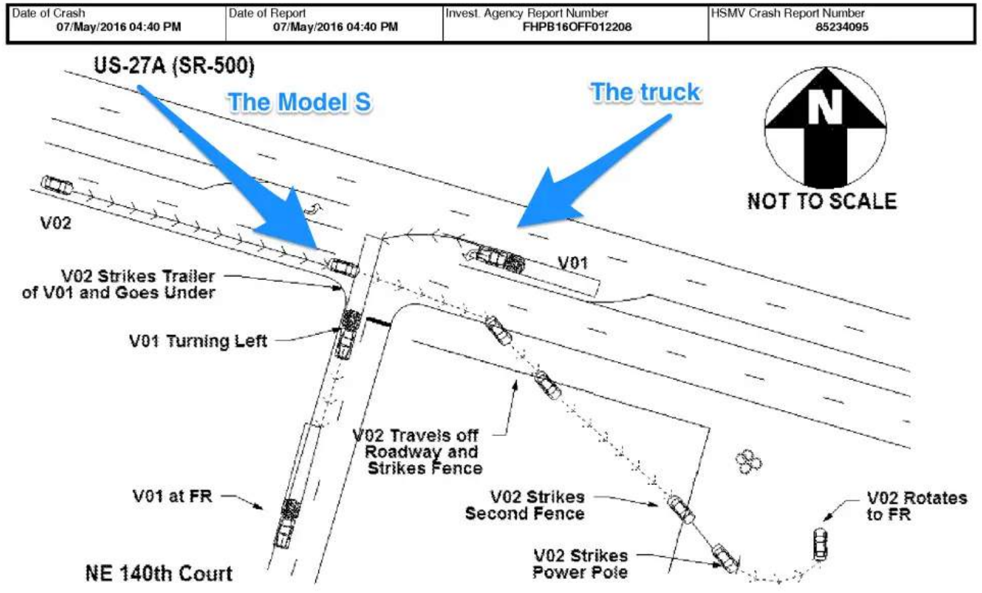


Disengagement: Cases where the car's software detects a failure or the driver perceived a failure, resulting in control being seized by the driver.

Source: https://www.statista.com/chart/17144/test-miles-and-reportable-miles-per-disengagement/

Tesla driver dies in first fatal crash while using autopilot mode

The autopilot sensors on the Model S failed to distinguish a white tractor-trailer crossing the highway against a bright sky



Autopilot didn't detect the trailer as an obstacle (NHTSA investigation and Tesla statements)

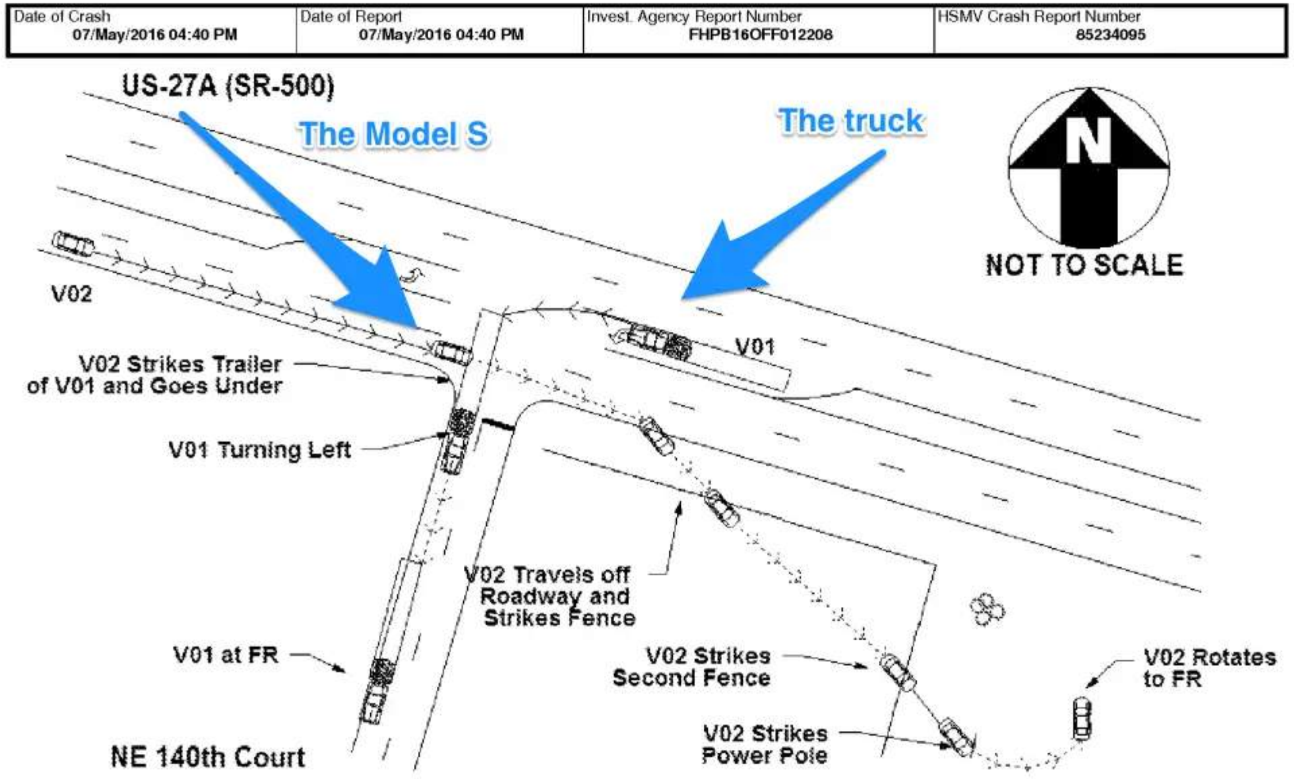
1. The National Highway Traffic Safety Administration (NHTSA) determined that a "lack of safeguards" contributed to the death
2. "Neither Autopilot nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied," Tesla said.

https://www.businessinsider.com/details-about-the-fatal-tesla-autopilot-accident-released-2017-6

Challenges in Perception in Autonomous Vehicles

Tesla driver dies in first fatal crash while using autopilot mode

1. The National Highway Traffic Safety Administration (NHTSA) determined that a “lack of safeguards” contributed to the death
2. "Neither Autopilot nor the driver noticed the white side of the tractor trailer against a brightly lit sky, so the brake was not applied," Tesla said.



https://www.businessinsider.com/details-about-the-fatal-tesla-autopilot-accident-released-2017-6

Perception in AVs
Setbacks and Challenges

Uber's self-driving SUV saw the pedestrian in fatal accident but didn't brake, officials say

PUBLISHED THU, MAY 24 2018•9:52 AM EDT | UPDATED THU, MAY 24 2018•10:43 AM EDT

Sensors on the fully autonomous Volvo XC-90 SUV spotted [redacted] while the car was traveling 43 miles per hour and determined that braking was needed 1.3 seconds before impact, according to the report.

1. https://www.telegraph.co.uk/technology/2018/03/20/ubers-fatal-accident-end-driverless-cars/
2. https://www.cnbc.com/2018/05/24/ubers-self-driving-suv-saw-the-pedestrian-in-fatal-accident-but-didnt-brake-officials-say.html

Perception in AVs
Technical Challenges

- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- Embedded perception
- V2X perception

Technical Challenges in Perception for AVs

Challenging Sensing and Weather

- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- Embedded perception
- V2X perception

Temel, Dogancan, et al. "Cure-tsd: Challenging unreal and real environments for traffic sign detection." *IEEE Transactions on Intelligent Transportation Systems* (2017).

Technical Challenges in Perception for AVs

Challenging Environments

- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- Embedded perception
- V2X perception

Dokania, S., Hafez, A. H., Subramanian, A., Chandraker, M., & Jawahar, C. V. (2023). IDD-3D: Indian Driving Dataset for 3D Unstructured Road Scenes. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 4482-4491).

Technical Challenges in Perception for AVs
Context Awareness

Does the fire impede mobility?

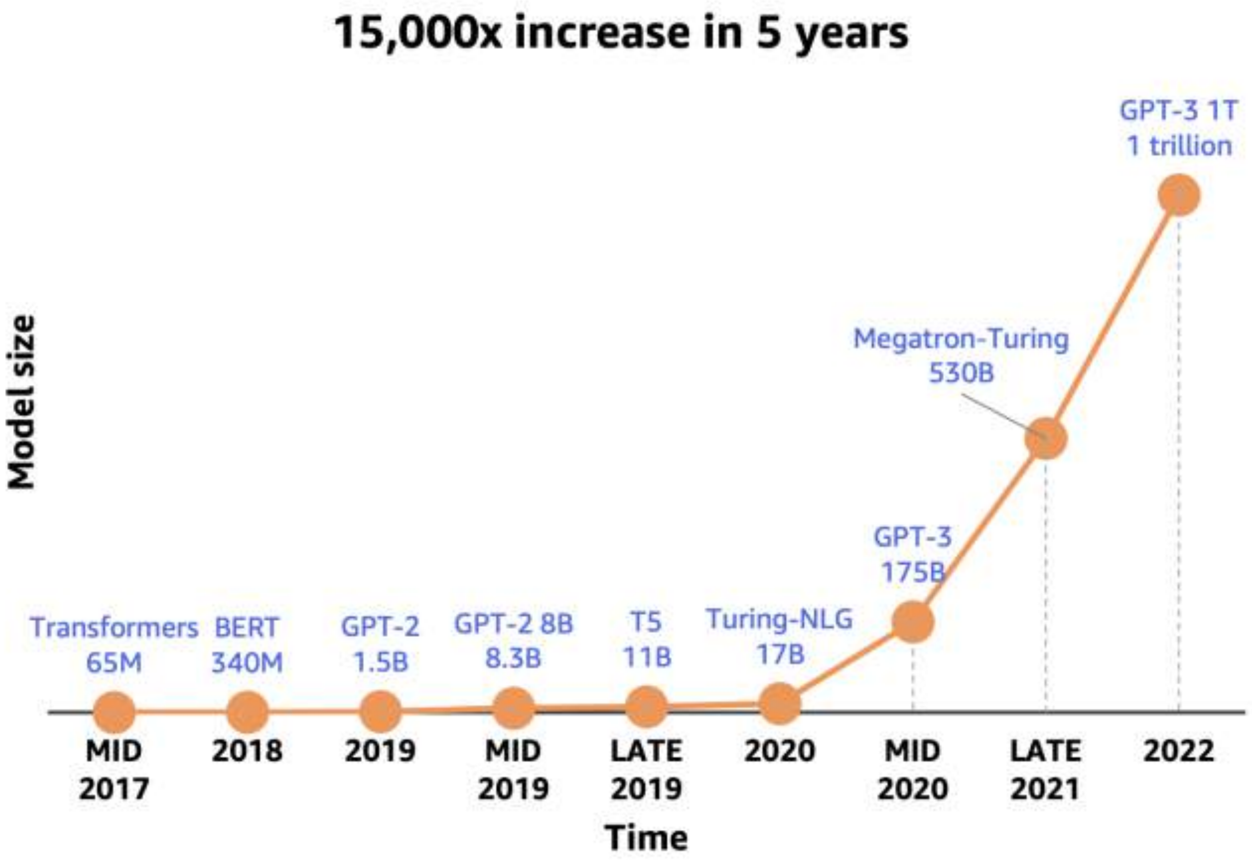
- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- Embedded perception
- V2X perception

Technical Challenges in Perception for AVs

Embedded Perception

On-board computational capabilities of modern deep learning algorithms is a challenge

- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- Embedded perception
- V2X perception



Technical Challenges in Perception for AVs
V2X Perception

Source: Fast and Furious 8!

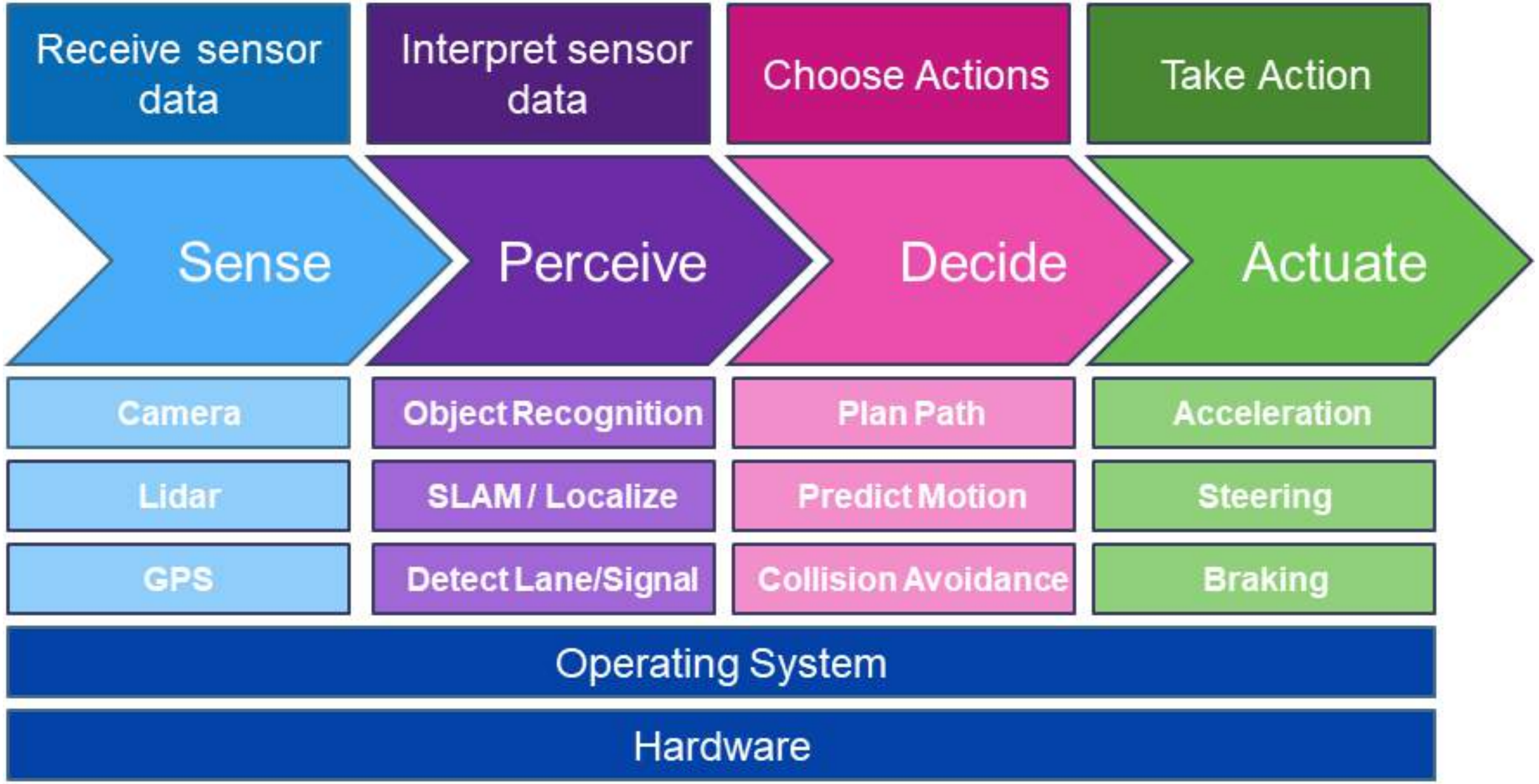
- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- Embedded perception
- V2X perception

Role of Perception
Role of Perception within AVs

Role of Perception:

- Filter,
- process, and
- understand

sensor data



Wards Intelligence, **Smarter Than Humans? AI for AVs: Sensing, Perception, Prediction and Planning**

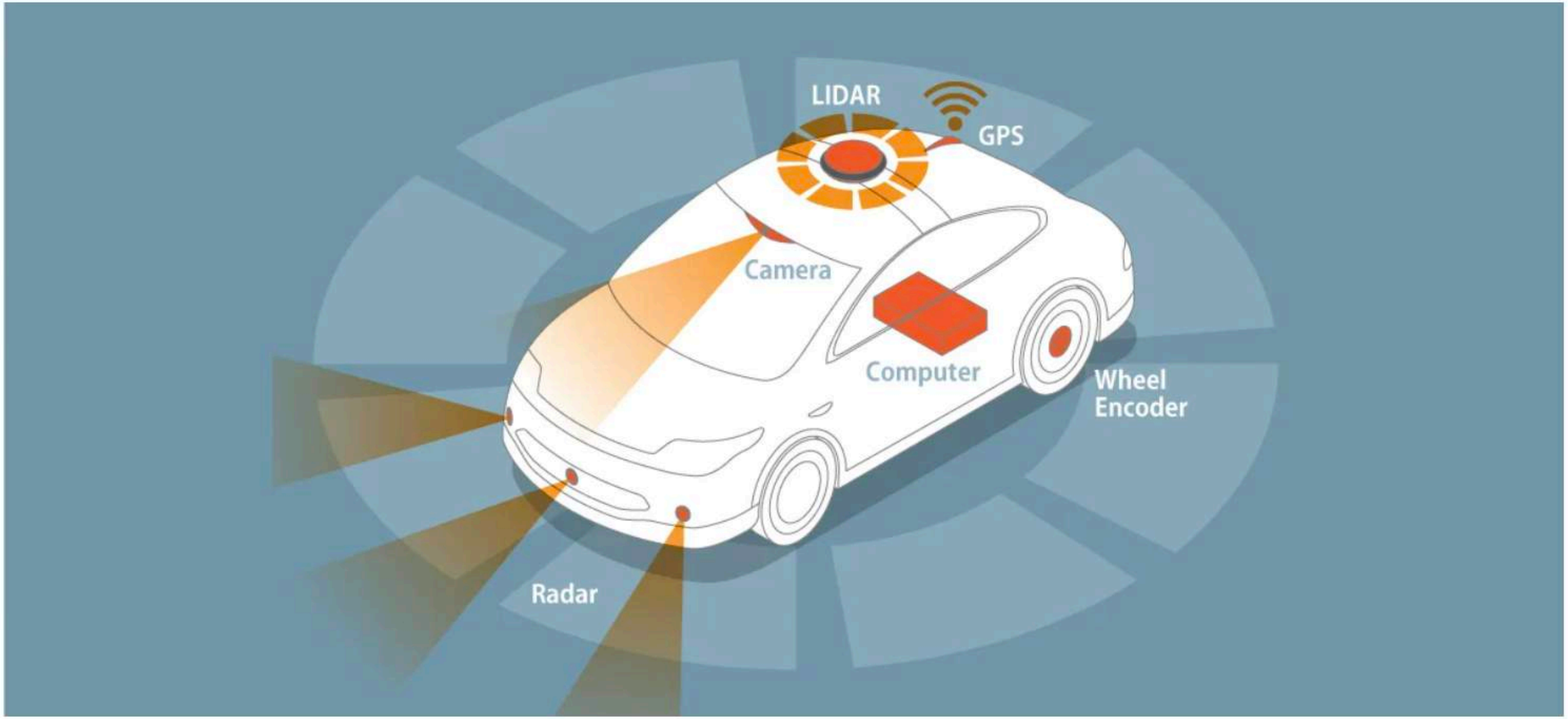
Sensors
Role of Sensors for Perception

Tsubaka Mechanical Engineering Laboratory (1977) | Eureka PROMETHEUS Project (1987 - 1995) | DARPA Grand Challenge (2004 - 2005)

More sensors and better fusion strategies!

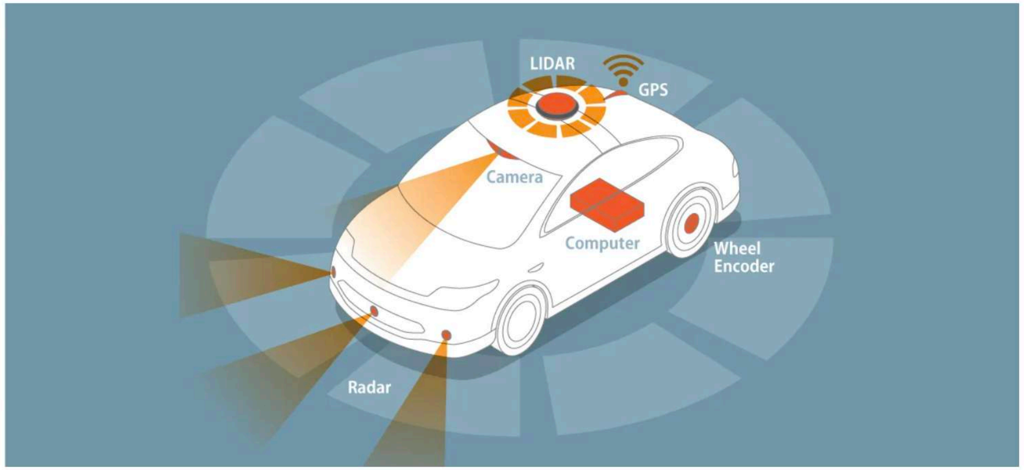
Sensors

How can we choose the “appropriate” Sensors?



Sensors
Choosing the Appropriate Sensors

- Sensors need to work under **challenging weather conditions**
- Sensors need to have sensing capacity and resolution in meeting **challenging sensing environments**
- Sensors must be **cost** effective
- Sensor fusion and sensor registration must be **computationally effective**
- Sensors must output minimum **noise** or their working ranges must be known in advance
- Sensor data must be resistant to **cyber and adversarial attacks**



Sensors
Choosing the Appropriate Sensors

Factors	Camera	LiDAR	Radar	Fusion
Range	~	~	✓	✓
Resolution	✓	~	×	✓
Distance Accuracy	~	✓	✓	✓
Velocity	~	×	✓	✓
Color Perception, e.g., traffic lights	✓	×	×	✓
Object Detection	~	✓	✓	✓
Object Classification	✓	~	×	✓
Lane Detection	✓	×	×	✓
Obstacle Edge Detection	✓	✓	×	✓
Illumination Conditions	×	✓	✓	✓
Weather Conditions	×	~	✓	✓

Yeong, D. J., Velasco-Hernandez, G., Barry, J., & Walsh, J. (2021). Sensor and sensor fusion technology in autonomous vehicles: A review. *Sensors*, *21*(6), 2140.

Sensors
Choosing the Appropriate Sensors

TABLE I
DIFFERENT SENSORS USED IN AV DEVELOPMENT

Vehicle	A#	B	C	D	E	F
Audi's Research Vehicle [48]	Y	Y	Y	Y	Y	Y
Ford: Hybrid Fusion [49]	Y			Y	Y	Y
Google: Toyota Prius [50]	Y	Y		Y	Y	
Nagoya and Nagasaki University's Open ZMP Robocar HV (Toyota Prius) [51]	Y			Y		
Volvo: (Stoklosa, Cars) [52]	Y		Y	Y	Y	Y
Apple: Lexus RX450h SUVs [53]	Y		Y	Y	Y	Y
DIDI's research vehicle [54]	Y		Y	Y	Y	Y
Infiniti Q50S [55]	Y				Y	Y
Lexus RX [56]	Y				Y	Y
Volvo XC90 [57]	Y				Y	Y
BMW750i xDrive [58]	Y	Y	Y		Y	Y
Mercedes-Benz E & S-Class [55]	Y	Y	Y		Y	Y
Otto Semi-Trucks [59]	Y			Y	Y	
Renault GT Nav [60]	Y				Y	Y
Tesla Model S [61]	Y				Y	Y
Baidu Apollo [62]	Y				Y	Y

#Note: A:Vision; B:Stereovision; C:IR Camera; D:LIDAR; E:Radar; and F:Sonar.

Ma, Y., Wang, Z., Yang, H., & Yang, L. (2020). Artificial intelligence applications in the development of autonomous vehicles: A survey. *IEEE/CAA Journal of Automatica Sinica*, *7*(2), 315-329.

Levels of Autonomy
Taxonomy

Current technology:

- Levels 1 and 2 are in the market
- Extensive testing on Level 3

SAE INTERNATIONAL

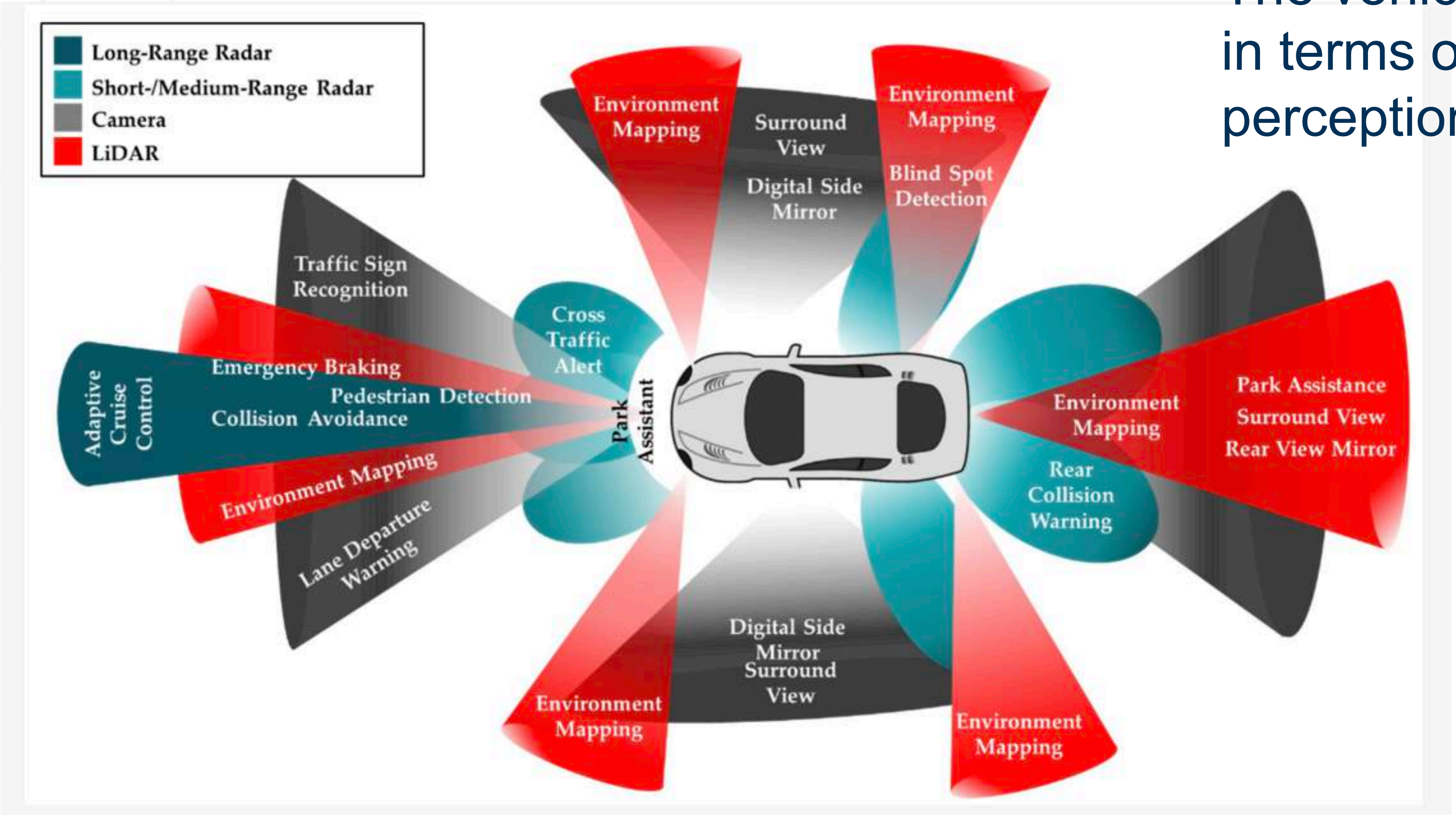
SAE J3016™ LEVELS OF DRIVING AUTOMATION™

Learn more here: sae.org/standards/content/j3016_202104

	SAE LEVEL 0™	SAE LEVEL 1™	SAE LEVEL 2™	SAE LEVEL 3™	SAE LEVEL 4™	SAE LEVEL 5™
What does the human in the driver's seat have to do?	**You <u>are</u> driving** whenever these driver support features are engaged – even if your feet are off the pedals and you are not steering			**You <u>are not</u> driving** when these automated driving features are engaged – even if you are seated in "the driver's seat"		
	You must constantly supervise these support features; you must steer, brake or accelerate as needed to maintain safety			When the feature requests, you must drive	These automated driving features will not require you to take over driving	
	These are driver support features			**These are automated driving features**		
What do these features do?	These features are limited to providing warnings and momentary assistance	These features provide steering **OR** brake/ acceleration support to the driver	These features provide steering **AND** brake/ acceleration support to the driver	These features can drive the vehicle under limited conditions and will not operate unless all required conditions are met		This feature can drive the vehicle under all conditions
Example Features	• automatic emergency braking • blind spot warning • lane departure warning	• lane centering **OR** • adaptive cruise control	• lane centering **AND** • adaptive cruise control at the same time	• traffic jam chauffeur	• local driverless taxi • pedals/ steering wheel may or may not be installed	• same as level 4, but feature can drive everywhere in all conditions

https://www.sae.org/blog/sae-j3016-update

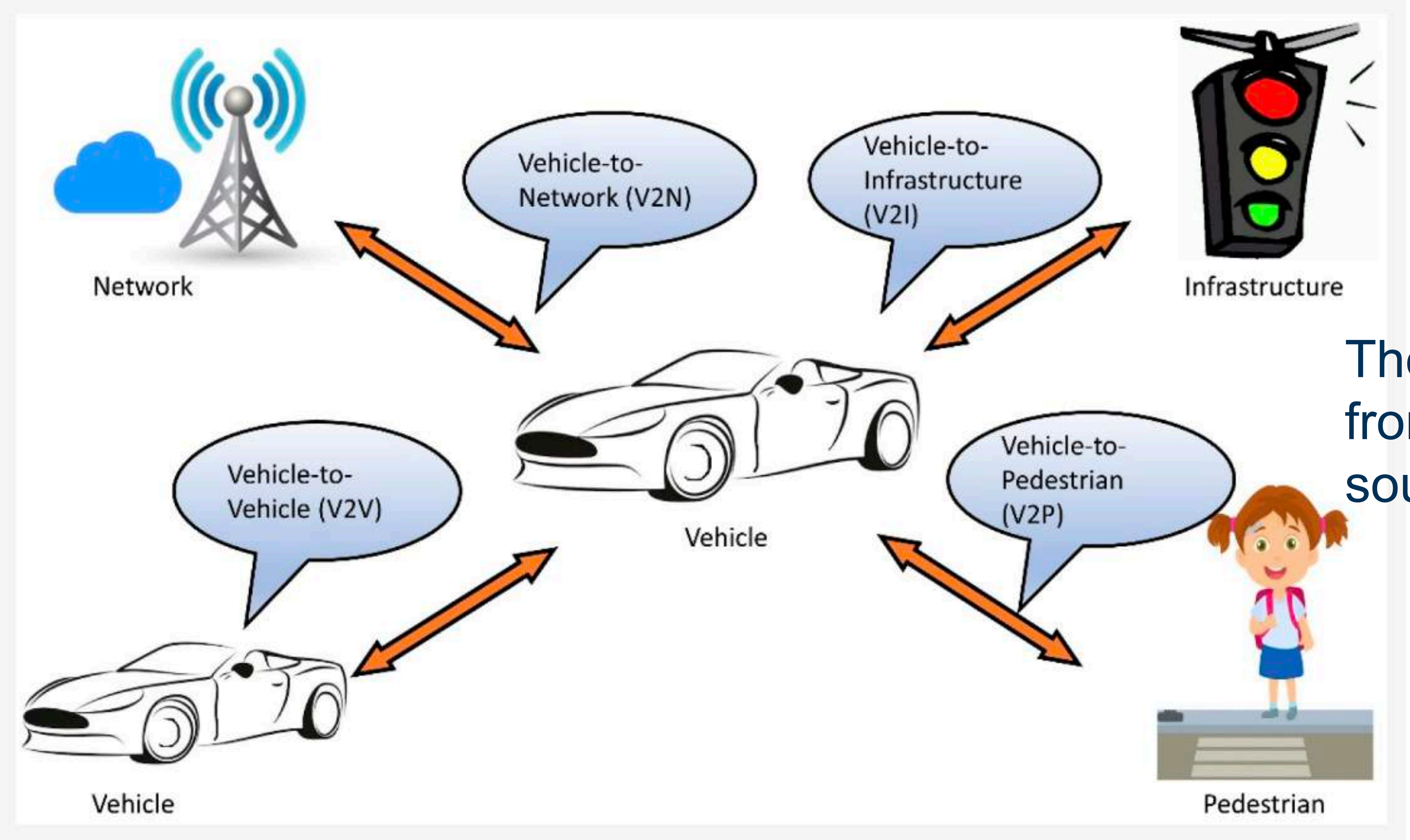
Levels of Autonomy
Levels 1 and 2 Autonomy



The vehicle is self-sufficient in terms of onboard sensors and perception!

Yeong, De Jong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. "Sensor and sensor fusion technology in autonomous vehicles: A review." *Sensors* 21, no. 6 (2021): 2140.

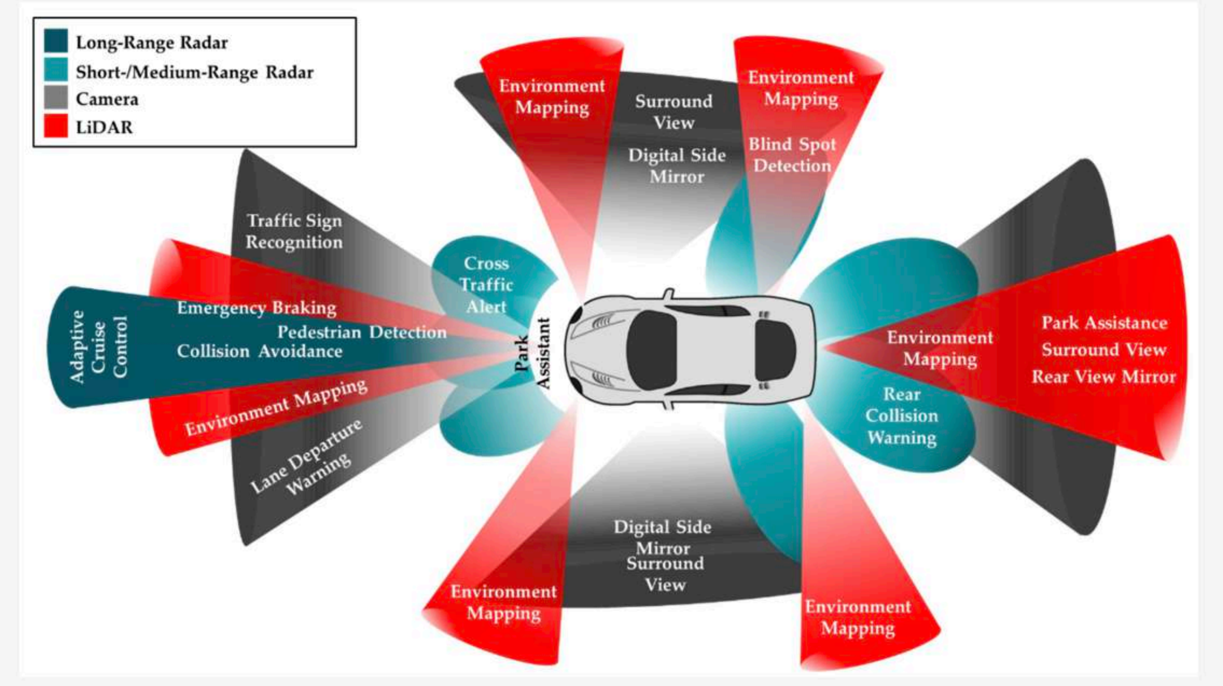
Levels of Autonomy
Levels 3 and Beyond

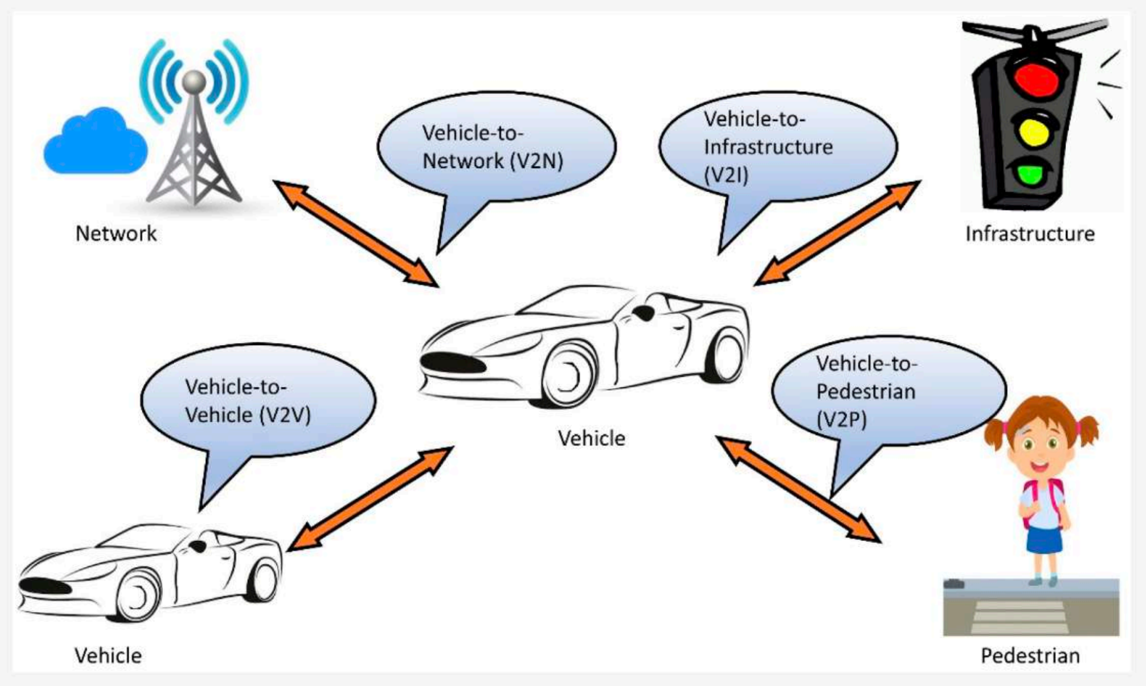


The vehicle needs help from other sensors, sources, and processors!

Haque, K. F., Abdelgawad, A., Yanambaka, V. P., & Yelamarthi, K. (2020). Lora architecture for v2x communication: An experimental evaluation with vehicles on the move. *Sensors*, *20*(23), 6876.

Levels of Autonomy
Achieving Perception



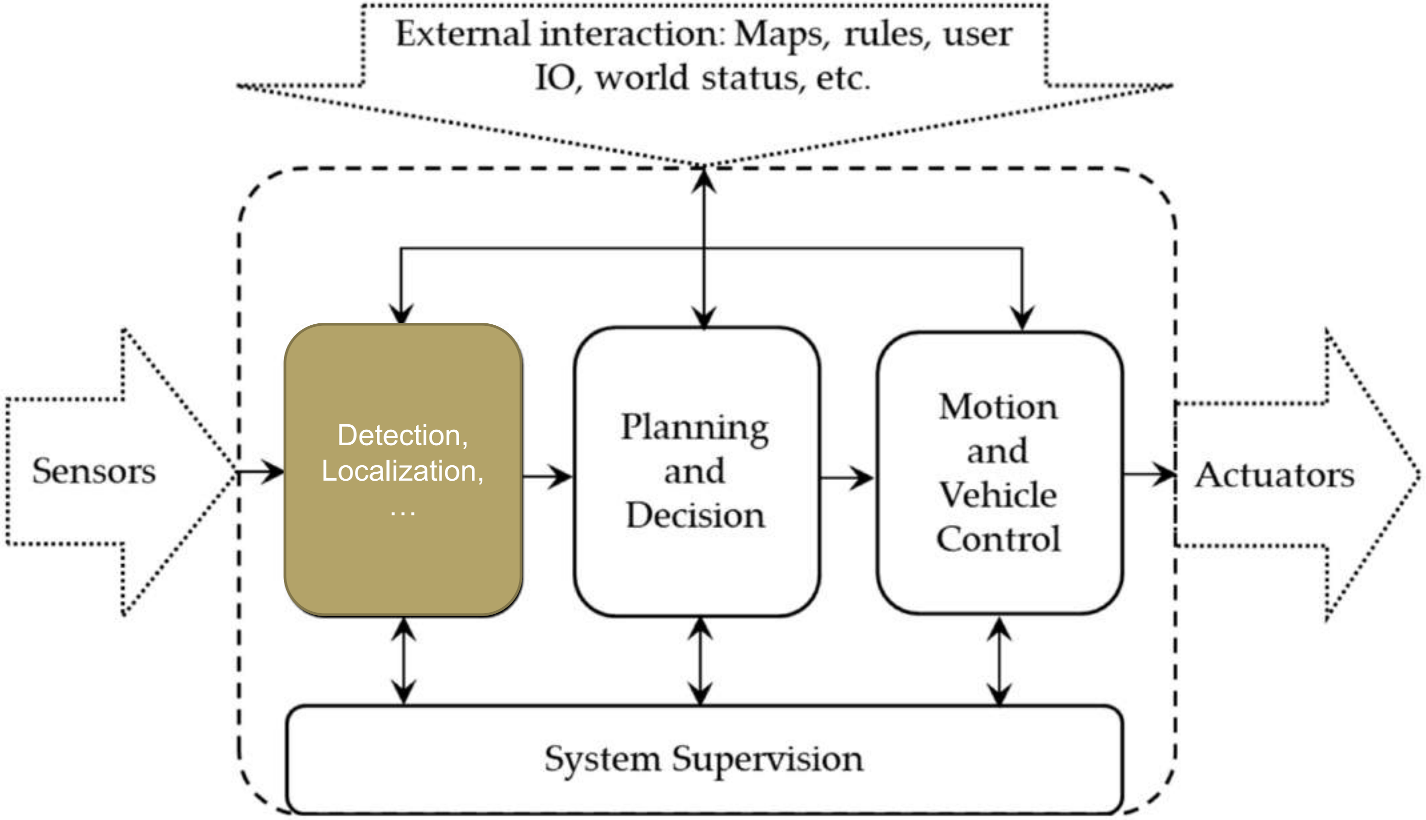


How to filter, process, and understand sensor data?

Yeong, De Jong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. "Sensor and sensor fusion technology in autonomous vehicles: A review." *Sensors* 21, no. 6 (2021): 2140.

Levels of Autonomy
Achieving Perception

Before: Perception is decomposed into a number of manageable applications

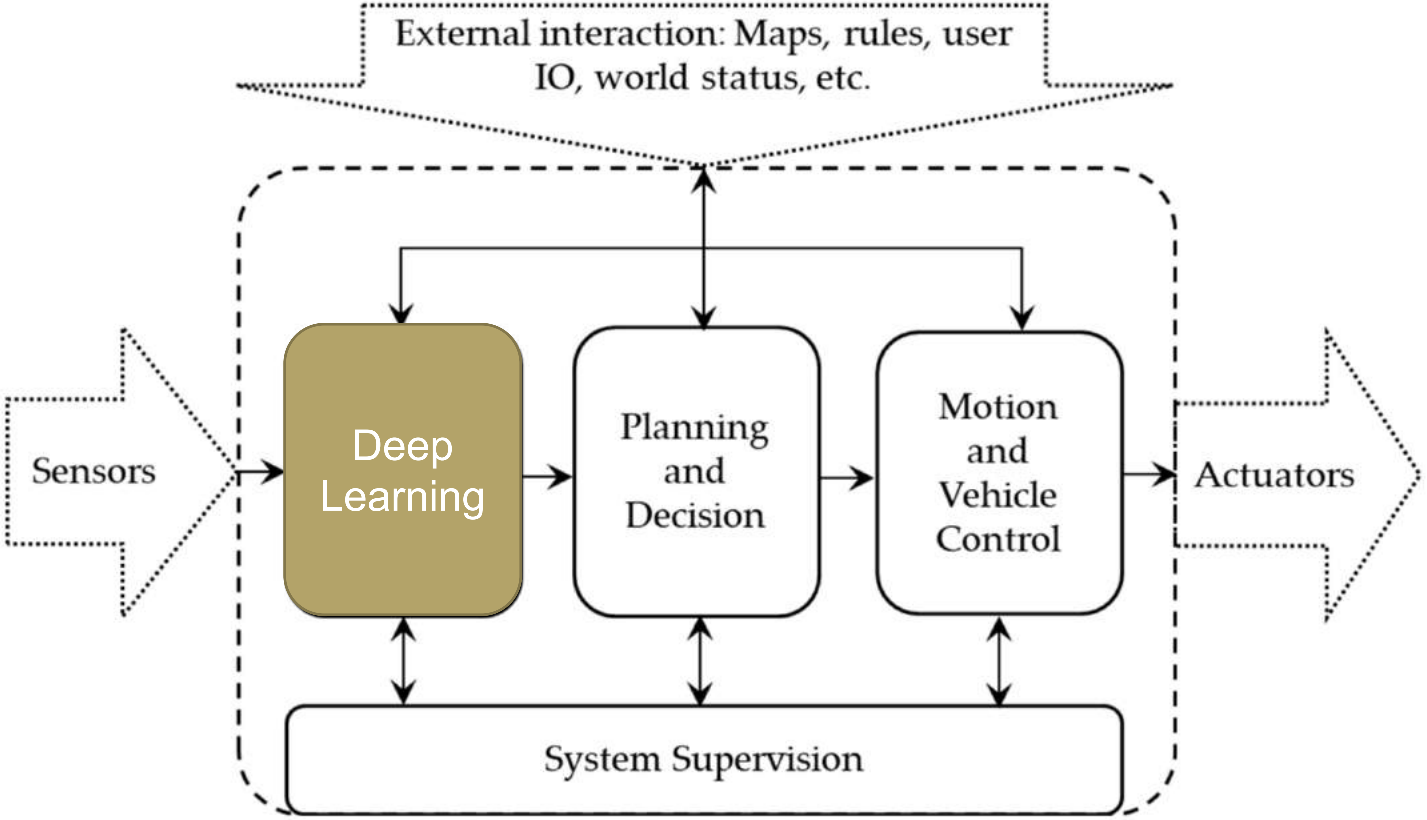


How to filter, process, and understand sensor data?

Yeong, De Jong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. "Sensor and sensor fusion technology in autonomous vehicles: A review." *Sensors* 21, no. 6 (2021): 2140.

Levels of Autonomy
Goal of the Tutorial

Deep Learning: Provides a holistic solution to perception



How to filter, process, and understand sensor data?

Yeong, De Jong, Gustavo Velasco-Hernandez, John Barry, and Joseph Walsh. "Sensor and sensor fusion technology in autonomous vehicles: A review." *Sensors* 21, no. 6 (2021): 2140.

Objectives

Takeaways from Part I

- **Part I: Challenges in Perception and Autonomy**
  - Robustness under challenging conditions, environments, context and surroundings-awareness are challenges in AV perception
  - Deep Learning promises a holistic solution to a number of the above challenges
- Part II: Deep Learning for Perception
- Part III: Existing Deep Learning solutions to Challenges in Perception
- Part IV: Remaining Challenges and Future Directions

A Holistic View of Perception in Intel. Vehicles

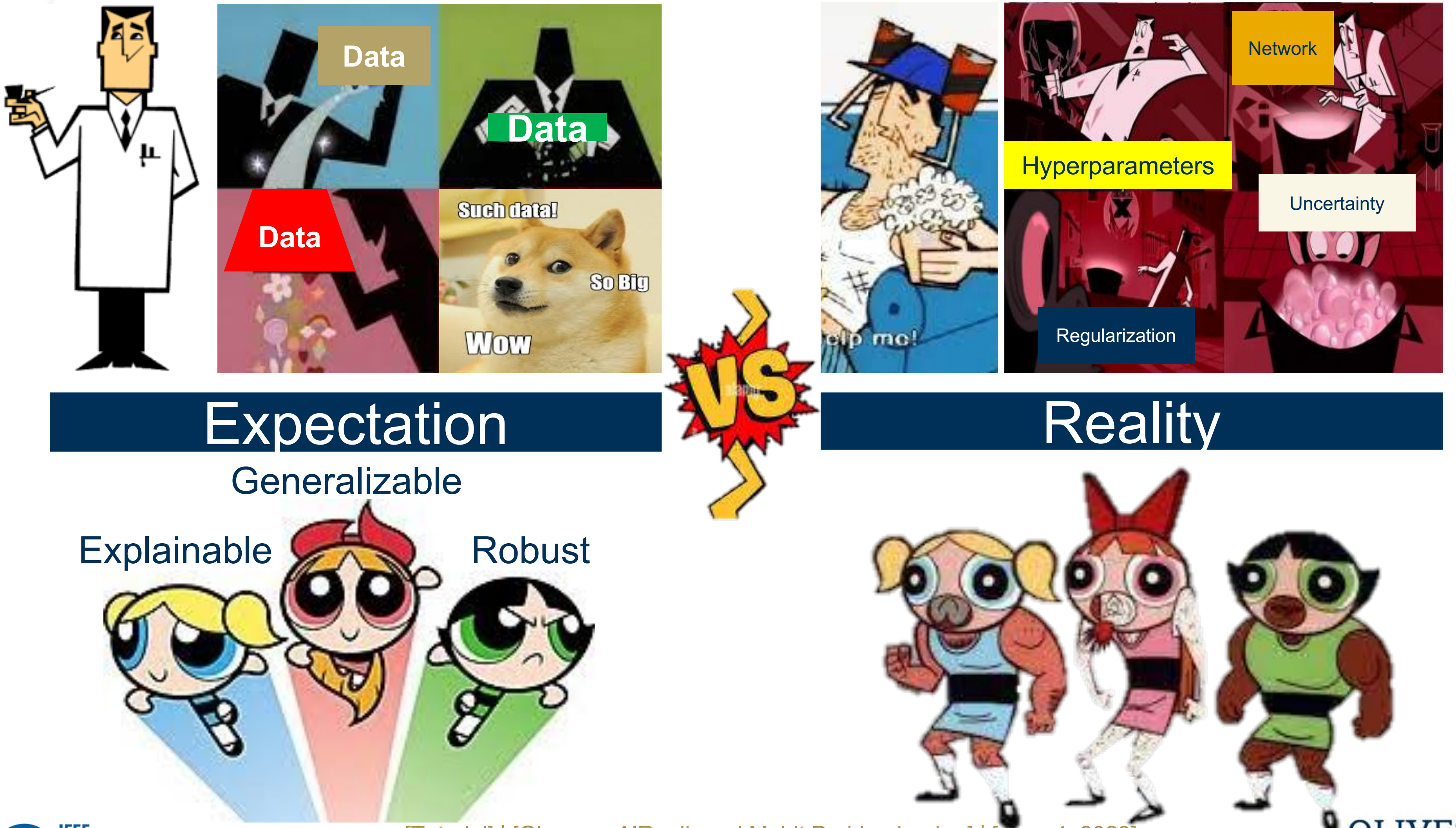
Part II: Deep Learning for Perception

Objectives

Objectives in Part II

- Discuss myths surrounding deep learning
- Brief history of deep learning
- Review deep learning models for vision
- Deep learning extensions into sensor domain
- Transfer Learning and foundation models
- Self-supervised learning
- Case study: Self-supervised learning for fisheye images

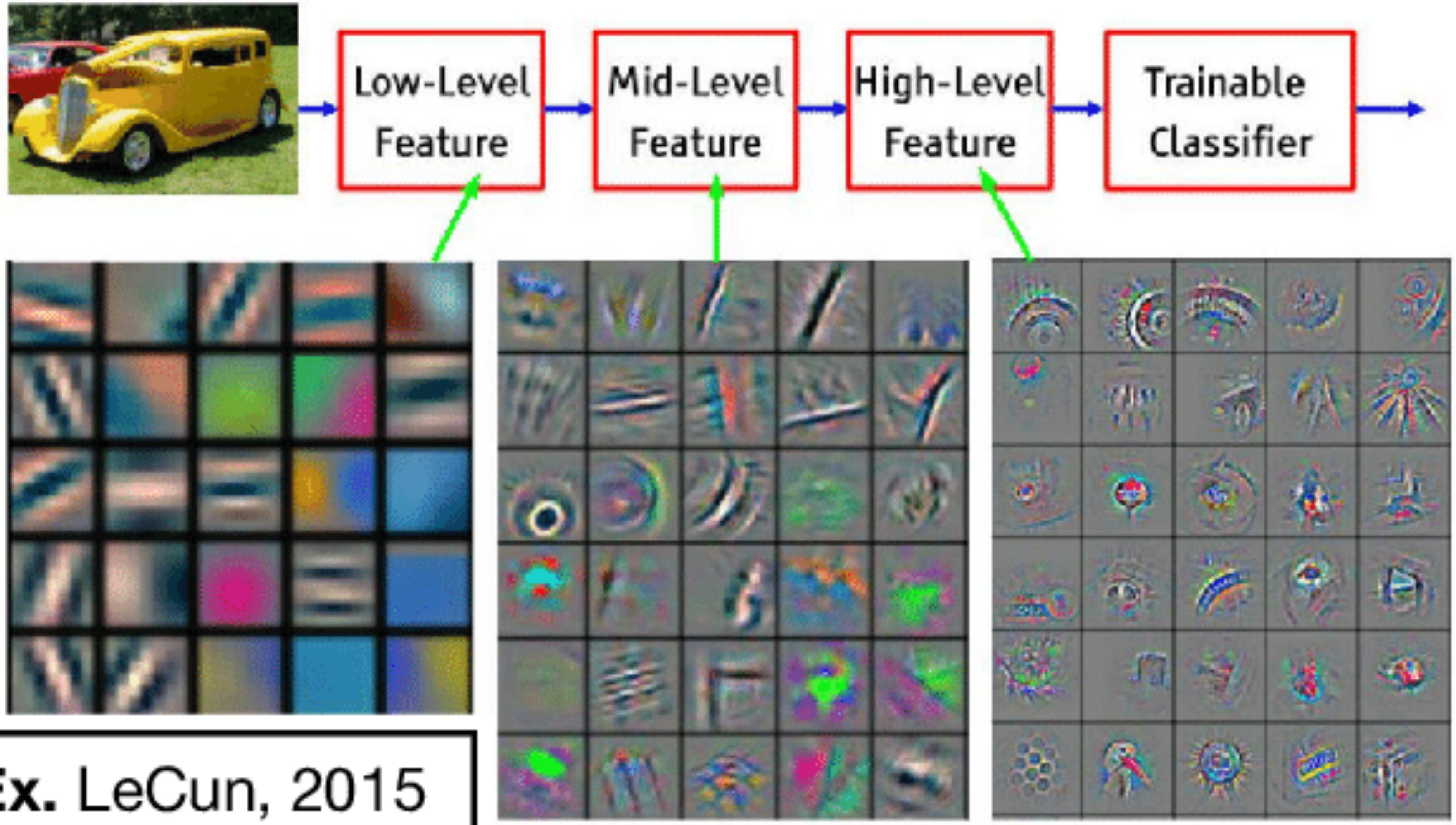
Deep Learning
Meme to start off with



Deep Learning
Meme to start off with

People’s expectation of AI and Deep Learning

Deep Learning
Model Decomposition



Ex. LeCun, 2015

“Deep learning is hard to train”

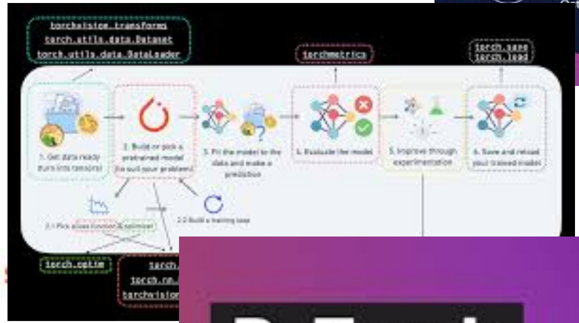
pytorch

109,392 repository results

Convolution Layers

nn.Conv1d	Applies a 1D convolution over an input signal composed of several input planes.
nn.Conv2d	Applies a 2D convolution over an input signal composed of several input planes.
nn.Conv3d	Applies a 3D convolution over an input signal composed of several input planes.
nn.ConvTranspose1d	Applies a 1D transposed convolution operator over an input image composed of several input planes.
nn.ConvTranspose2d	Applies a 2D transposed convolution operator over an

- Containers
- Convolution Layers
- Pooling layers
- Padding Layers
- Non-linear Activations (weighted
- Non-linear Activations (other)
- Normalization Layers
- Recurrent Layers
- Transformer Layers
- Linear Layers

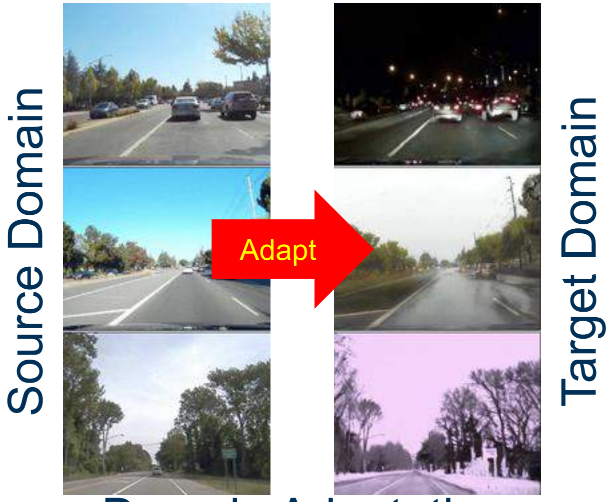


Deep Learning

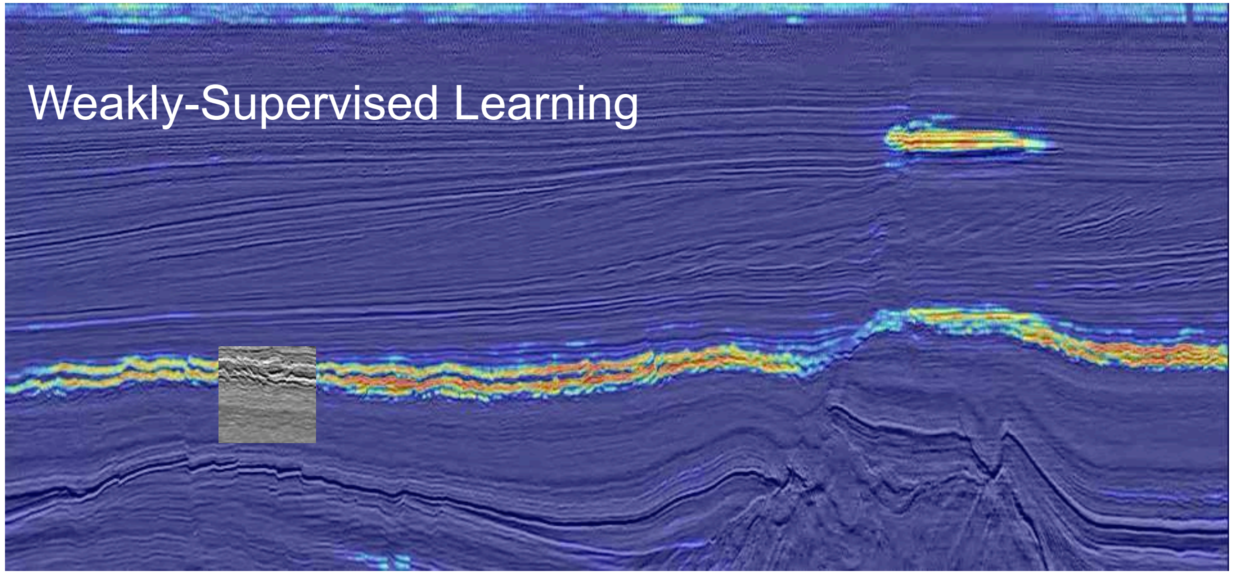
Some Common Myths about Deep Learning

“Deep learning requires lots of data”

Zero-Shot Learning

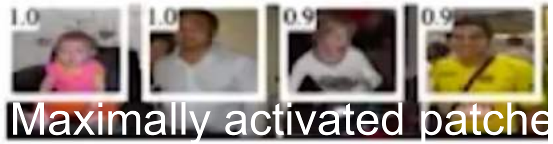


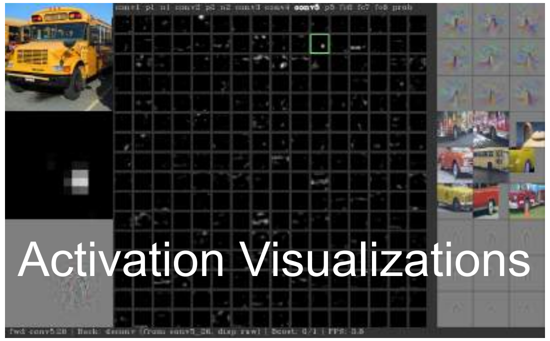
Domain Adaptation

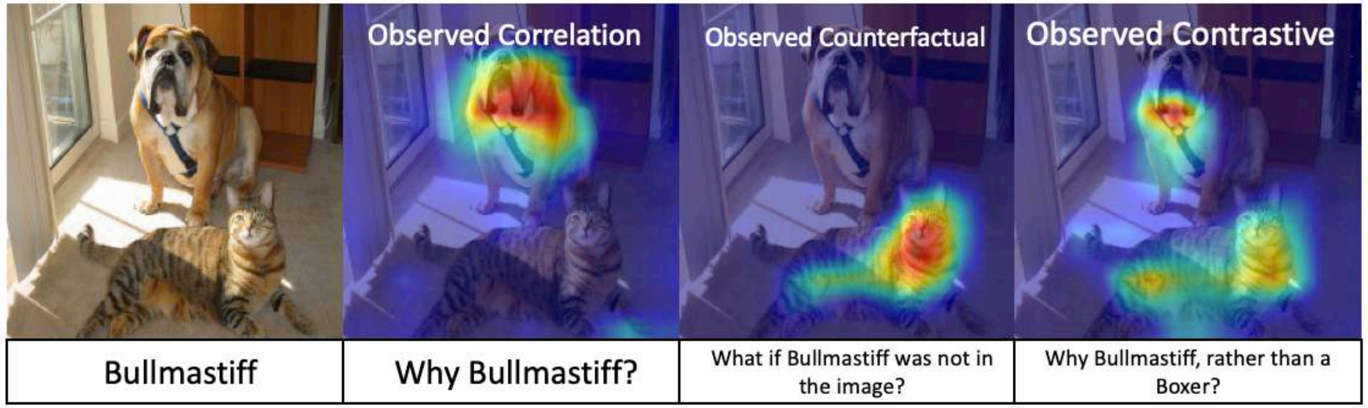


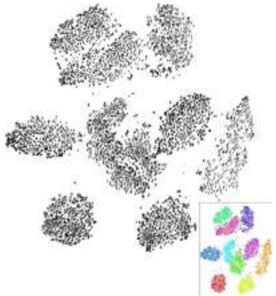
Deep Learning
Some Common Myths about Deep Learning

"Deep learning has poor interpretability"

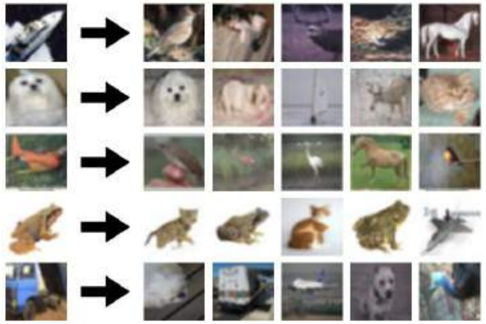






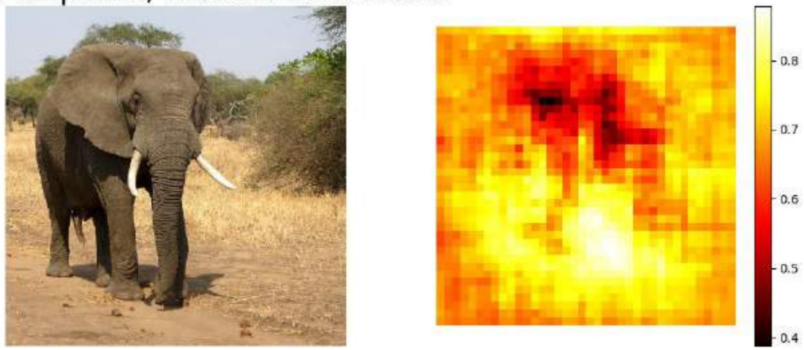


Dimensionality Reduction



Nearest Neighbor

African elephant, Loxodonta africana

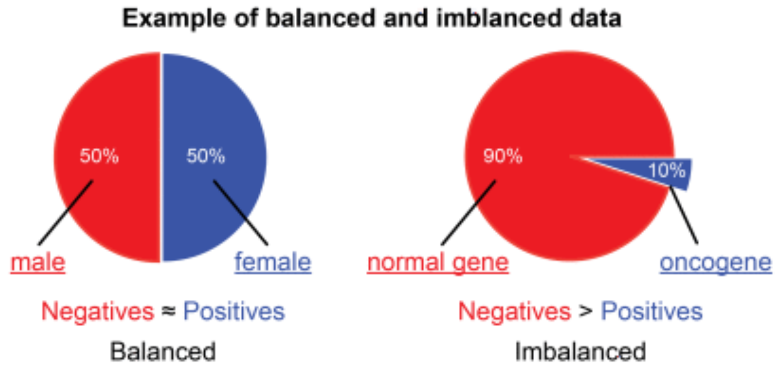


Saliency via occlusion

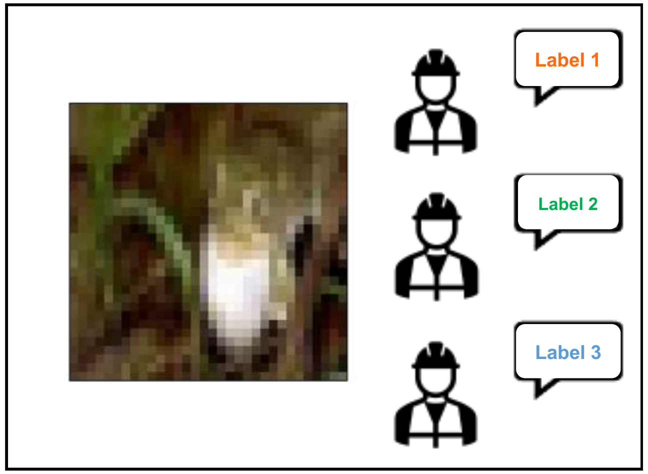
Deep Learning

Some Common Myths about Deep Learning

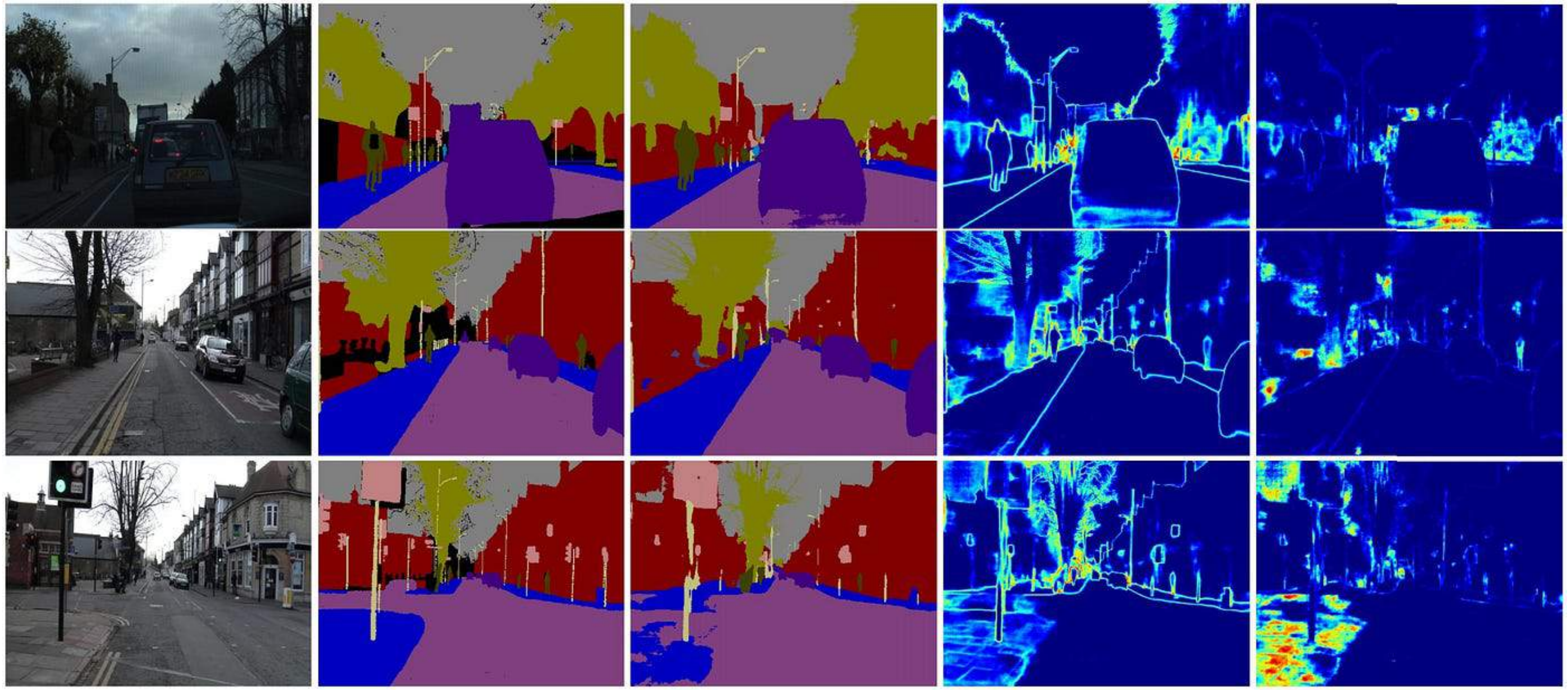
“More the data, better the model”



Data imbalance issues



Human labeling issues



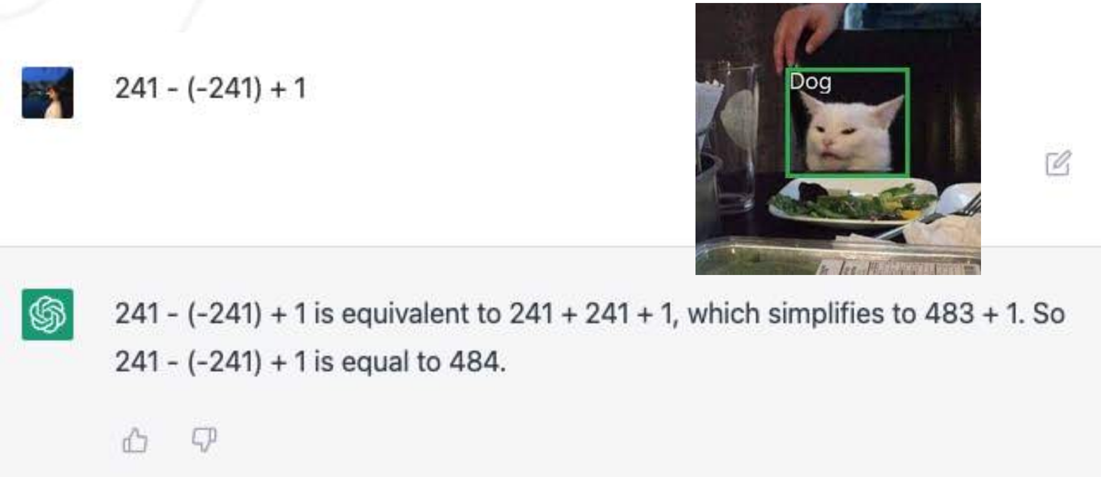
(a) Input Image (b) Ground Truth (c) Semantic Segmentation (d) Aleatoric Uncertainty (e) Epistemic Uncertainty

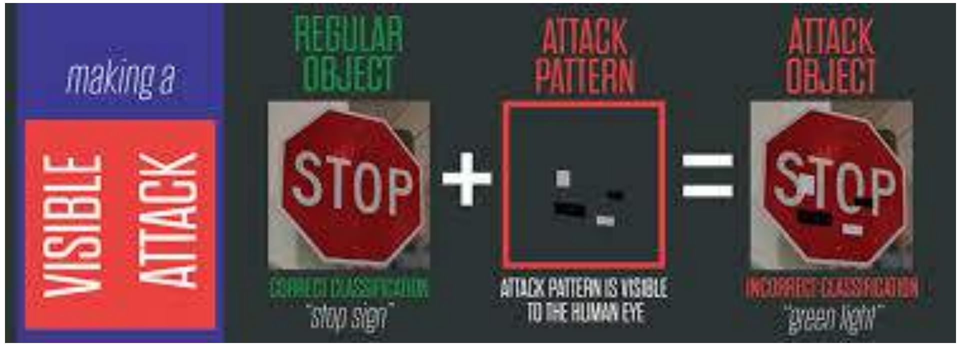
Dataset uncertainties

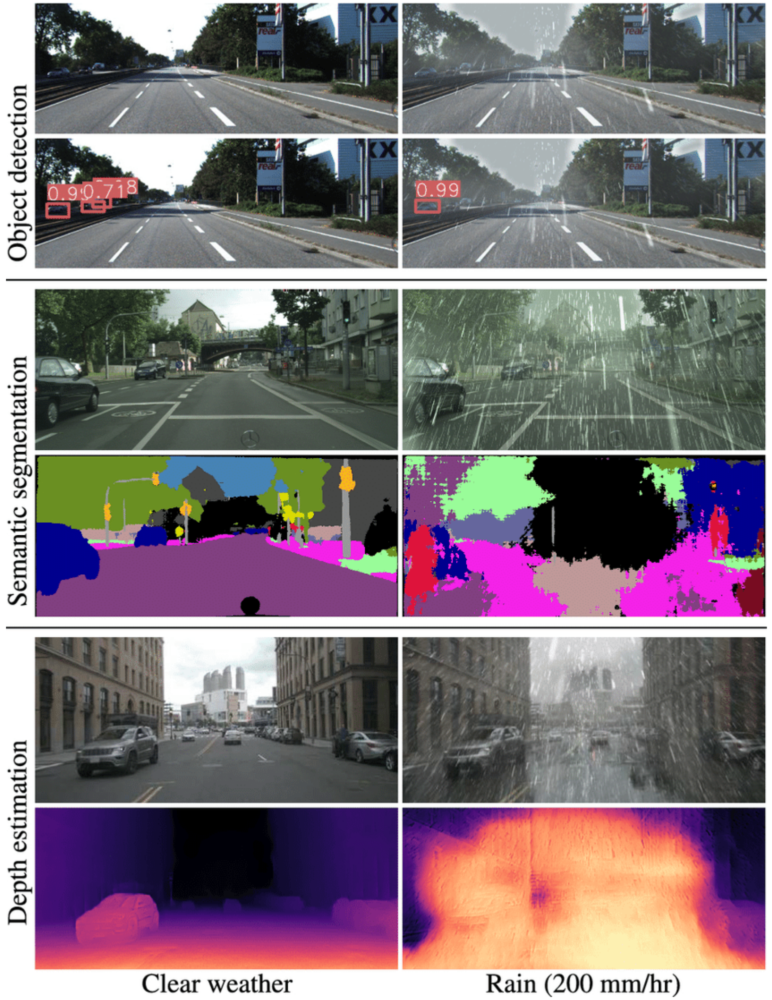
Deep Learning

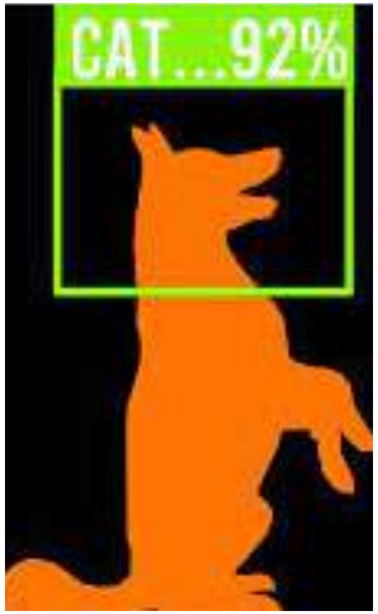
Some Common Myths about Deep Learning

"Deep learning is State-of-the-Art in every field"





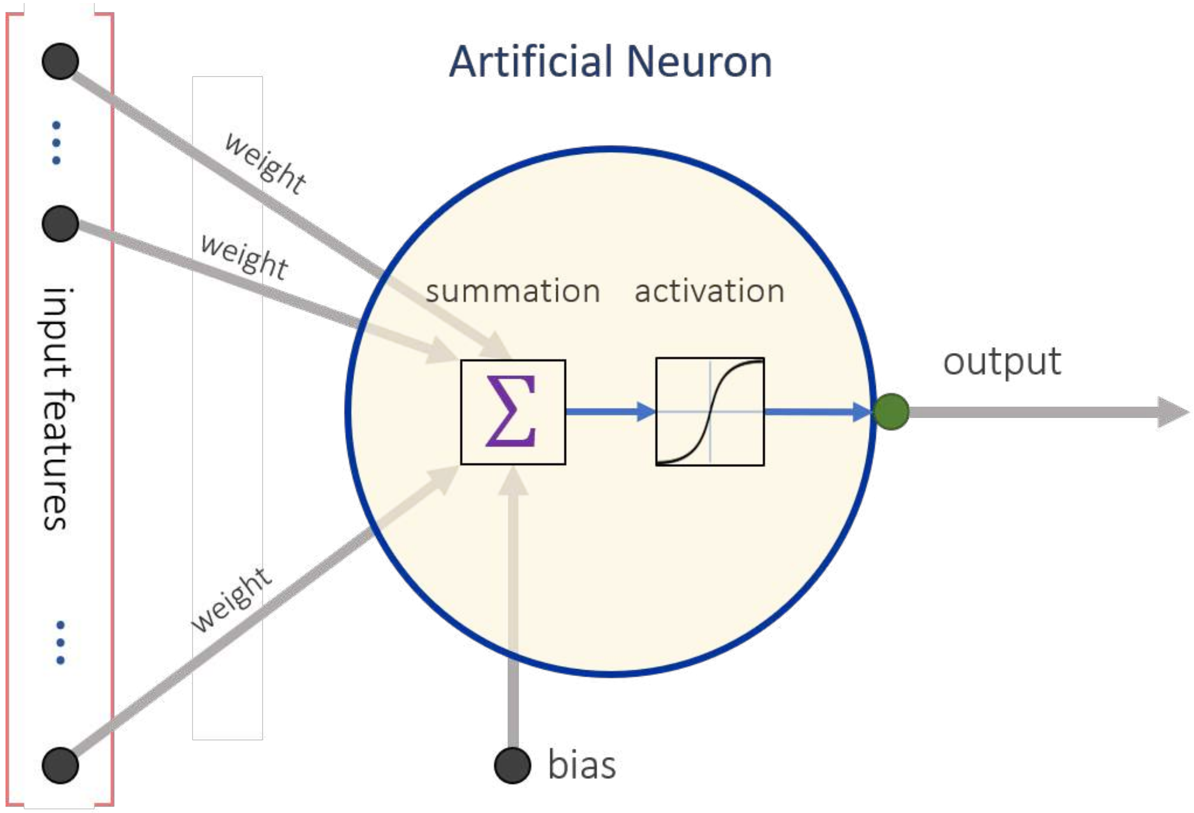




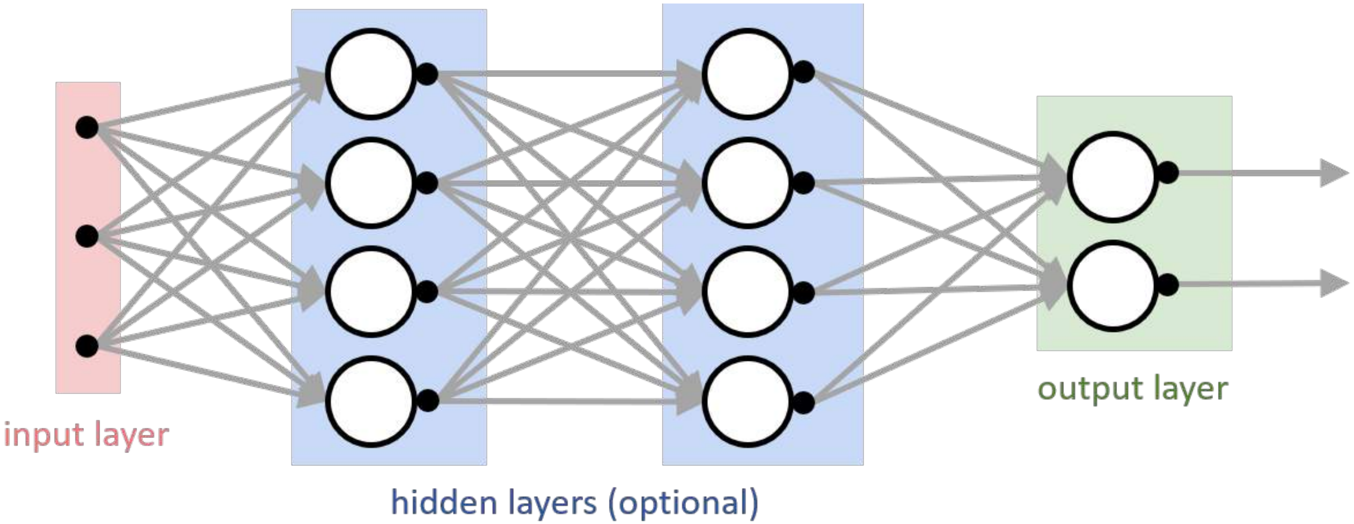
The underlying computational unit is the artificial neuron

Artificial neurons consist of:

- A single **output**
- Multiple **inputs**
- Input **weights**
- A **bias** input
- An **activation function**



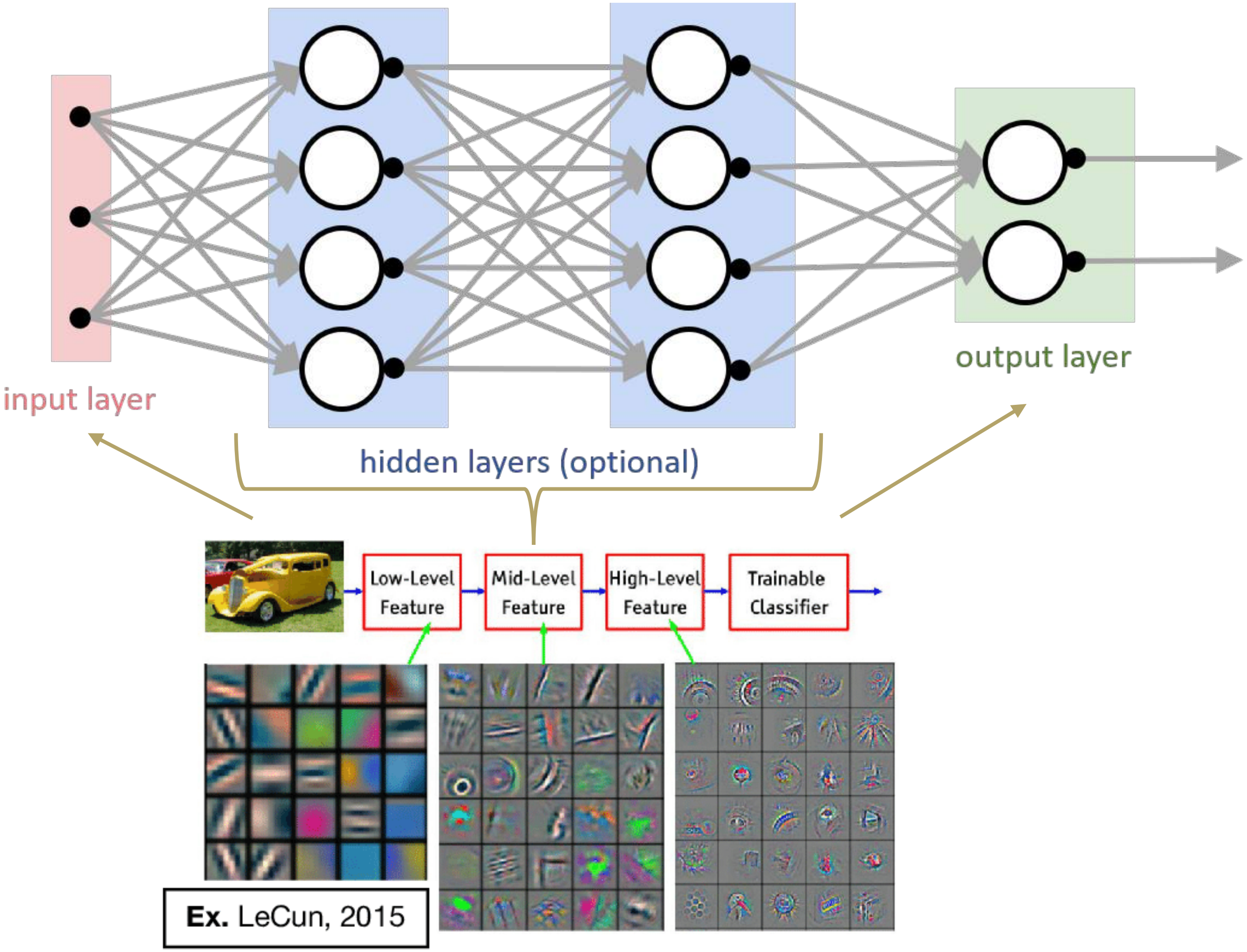
Deep Learning
Artificial Neural Networks



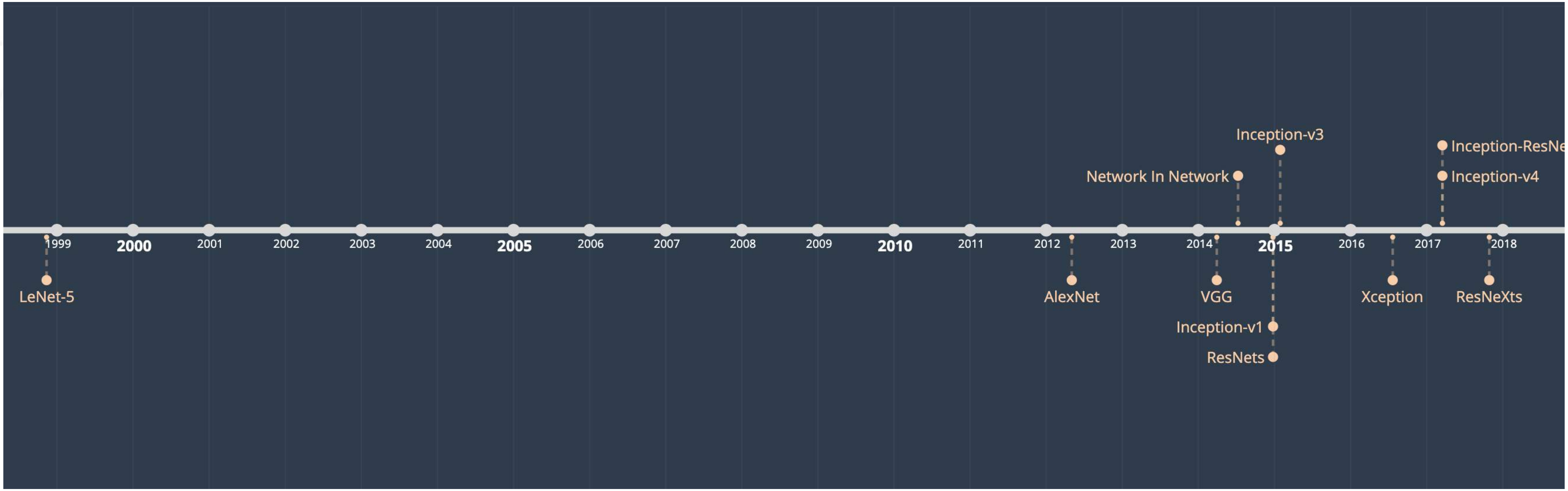
Typically, a neuron is part of a network organized in **layers**:

- An **input** layer (Layer 0)
- An **output** layer (Layer K)
- Zero or more **hidden** (middle) layers (Layers $1 \ldots K-1$)

Deep Learning
Convolutional Neural Networks

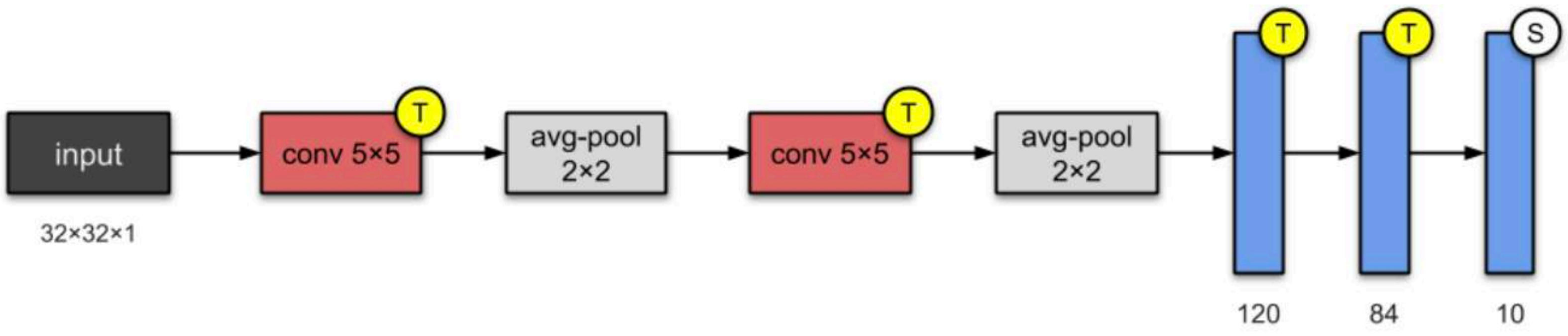


- LeNet
- AlexNet
- VGG
- GoogLeNet (Inception-V1)
- ResNet



CNN Architectures

LeNet5 (1998)



Novelty:

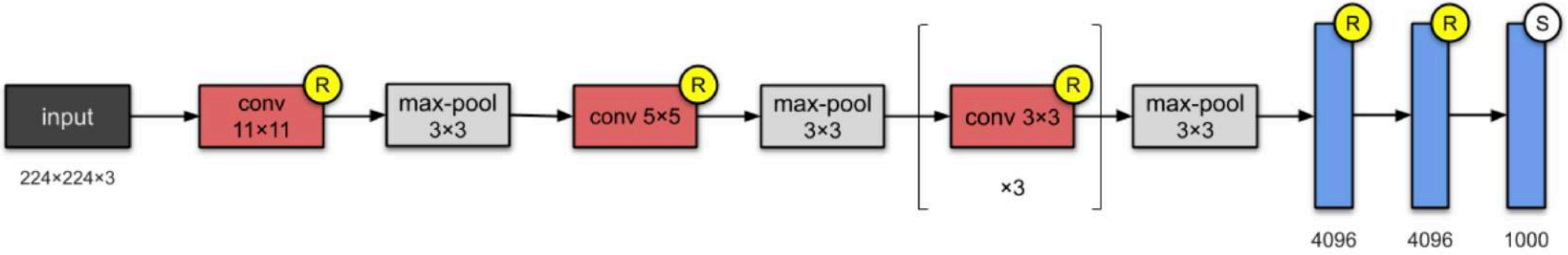
- Reduced number of learnable parameters and learned from raw pixels automatically
- The 1st popular CNN that became the "standard" template of CNNs
  - Stacking convolutional, activation, pooling layers
  - Ending with fully connected layers
- Good results on small datasets
  - Top-5 error rate on MNIST is 0.95%

Long Gap (1998 – 2012)

- Working to improve **computational power**
  - Existing accelerators were not yet sufficiently powerful to make deep multichannel, multilayer CNNs with a large number of parameters.
- Existing **datasets** were relatively **small**
  - Limited storage capacity of computers
- **Tricks** for **neural network training** were not established yet
  - Parameter initialization
  - Variants of stochastic gradient descent
  - Non-squashing activation functions
  - Effective regularization techniques

https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d

CNN Architectures
AlexNet (2011)

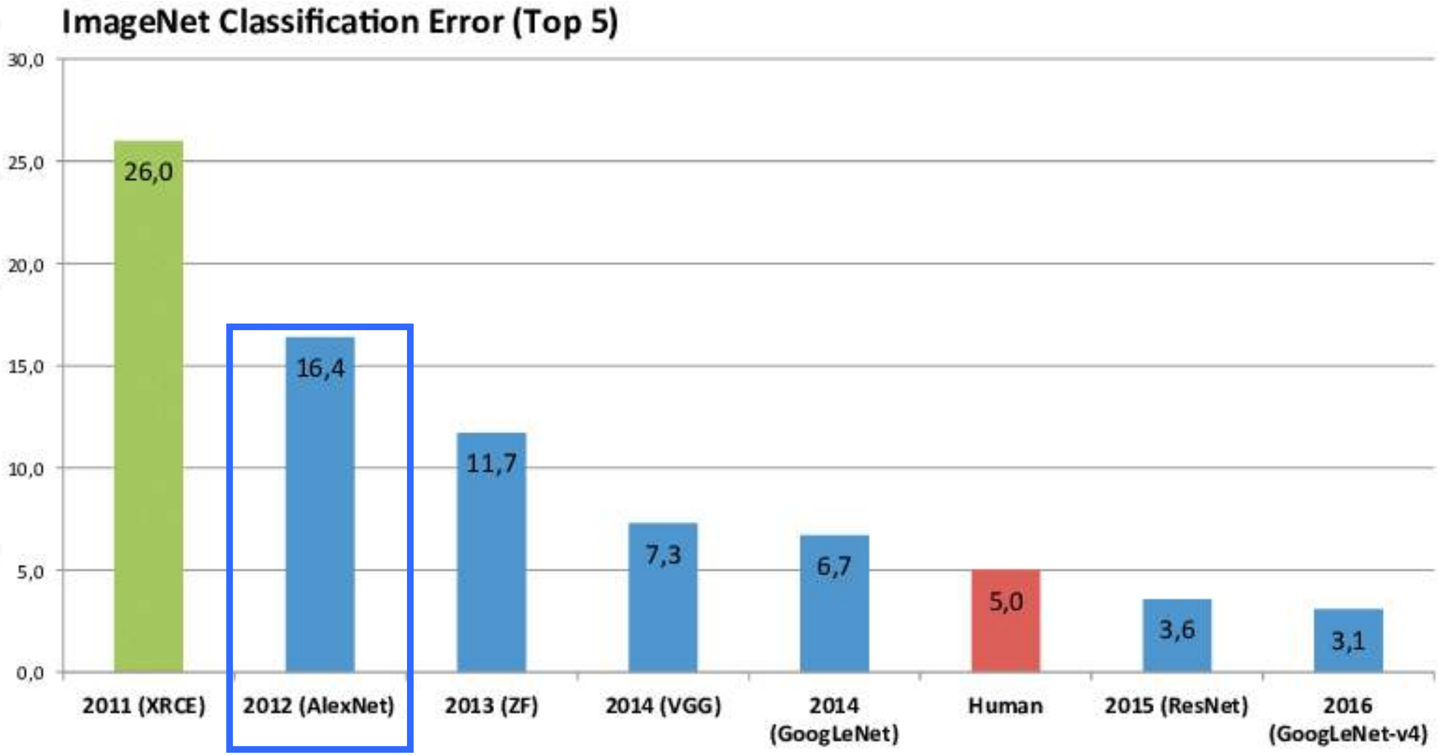


Novelty:

- First to implement Rectified Linear Units (ReLUs) as activation, solving the vanishing gradient problem
- Applied dropout regularization to fully connected layer to control complexity
- Deep CNN that runs on GPU hardware
- Deeper and wider than LeNet
- More robust than LeNet (data augmentation)
- **Won ImageNet Challenge and significantly outperformed traditional methods**

https://towardsdatascience.com/illustrated-10-cnn-architectures-95d78ace614d

AlexNet (2012)



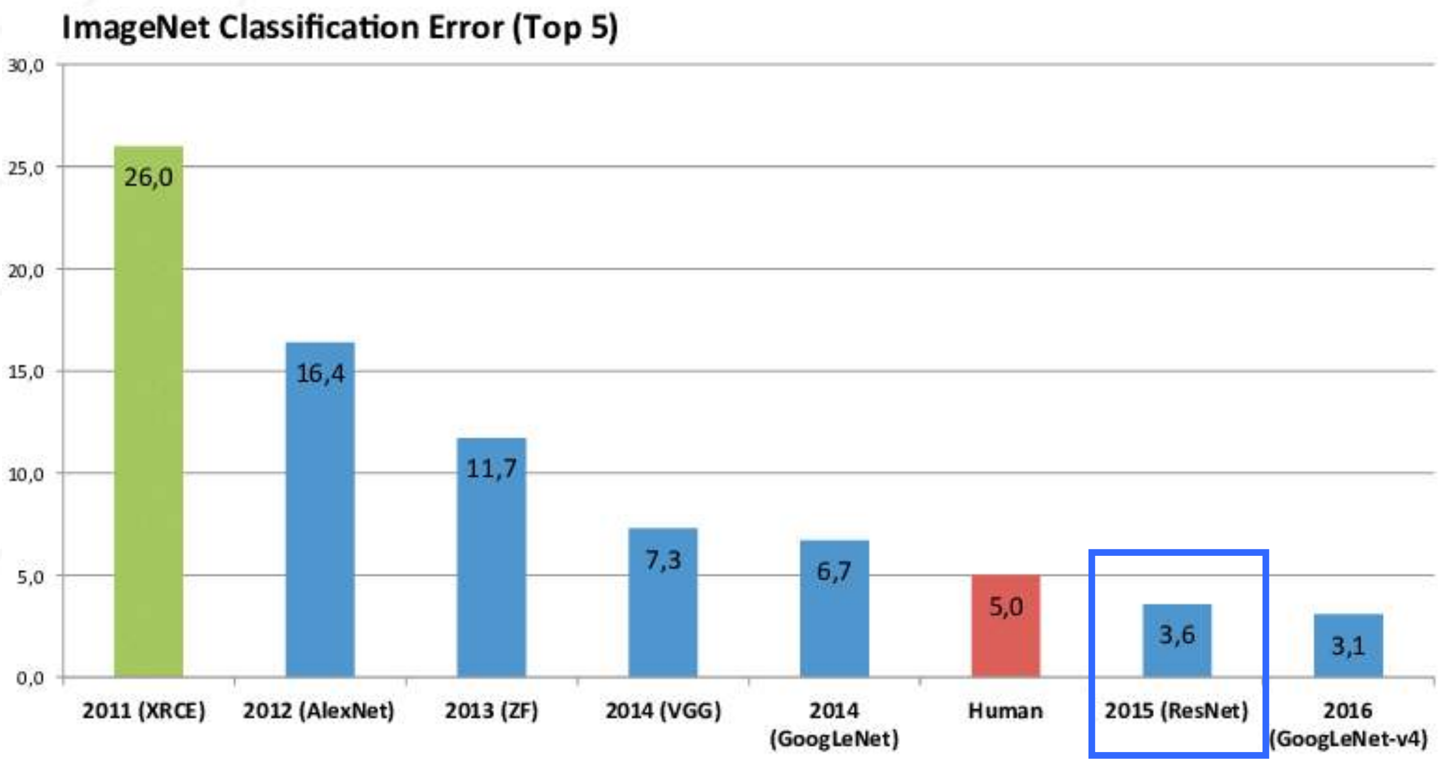
Imagenet:
1000 classes, 1.2M training images, 150K for testing

16.4% top 5 error in ILSVRC 2012

Figure Credit: Zitzewitz, Gustav. "Survey of neural networks in autonomous driving." (2017)

Krizhevsky et al. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012)

ResNet (2015)



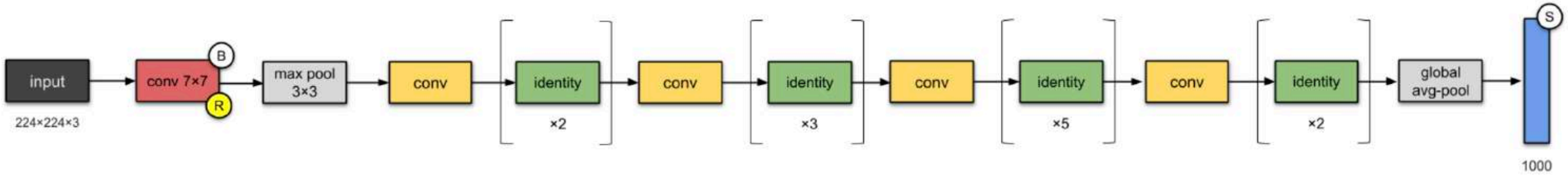
~3.6% top 5 error in ILSVRC 2015,
lower than human recognition error!

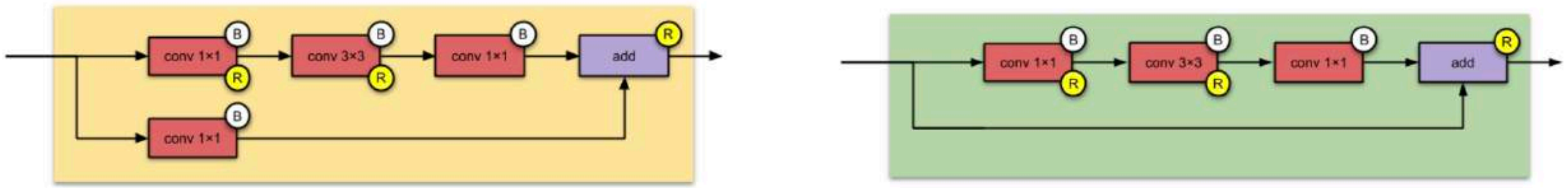
Imagenet:
1000 classes, 1.2M training images, 150K for testing

Figure Credit: Zitzewitz, Gustav. "Survey of neural networks in autonomous driving." (2017)

He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

CNN Architectures
ResNet (2015)



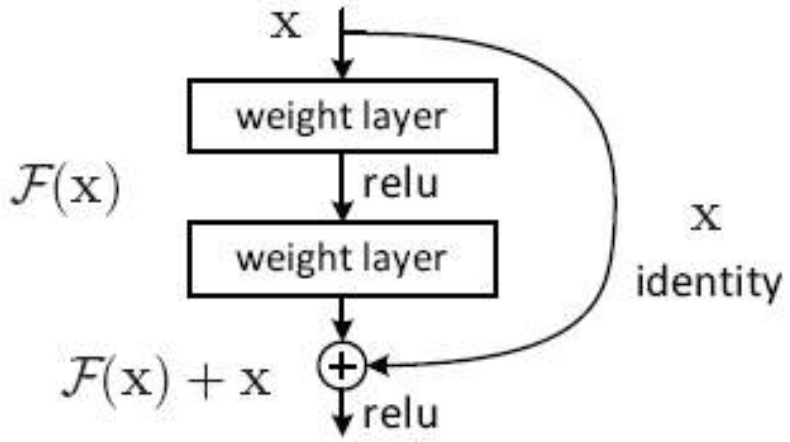


Conv block

Identity block

Novelty:

- Introduced residual learning (Residual blocks)
  - Shortcut connections with identity mapping
- Popularized skip connections
- 20 and 8 times deeper than AlexNet and VGG, respectively with less computational complexity and without compromising generalization power



[Tutorial] | [Ghassan AlRegib and Mohit Prabhushankar] | [June 4, 2023]

He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.

Object Detection Architectures
YOLO (2016 - Ongoing)

All previous object detection techniques required multiple stages of detection

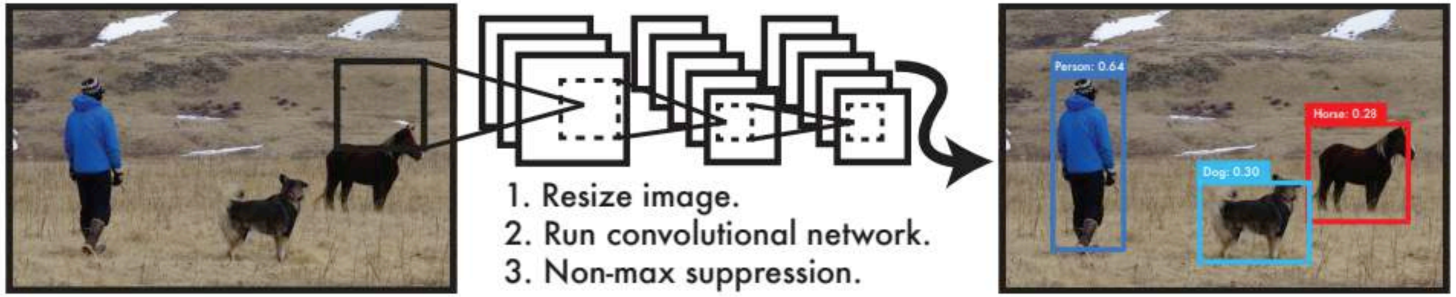


Figure 1: The YOLO Detection System. Processing images with YOLO is simple and straightforward. Our system (1) resizes the input image to 448 × 448, (2) runs a single convolutional network on the image, and (3) thresholds the resulting detections by the model's confidence.

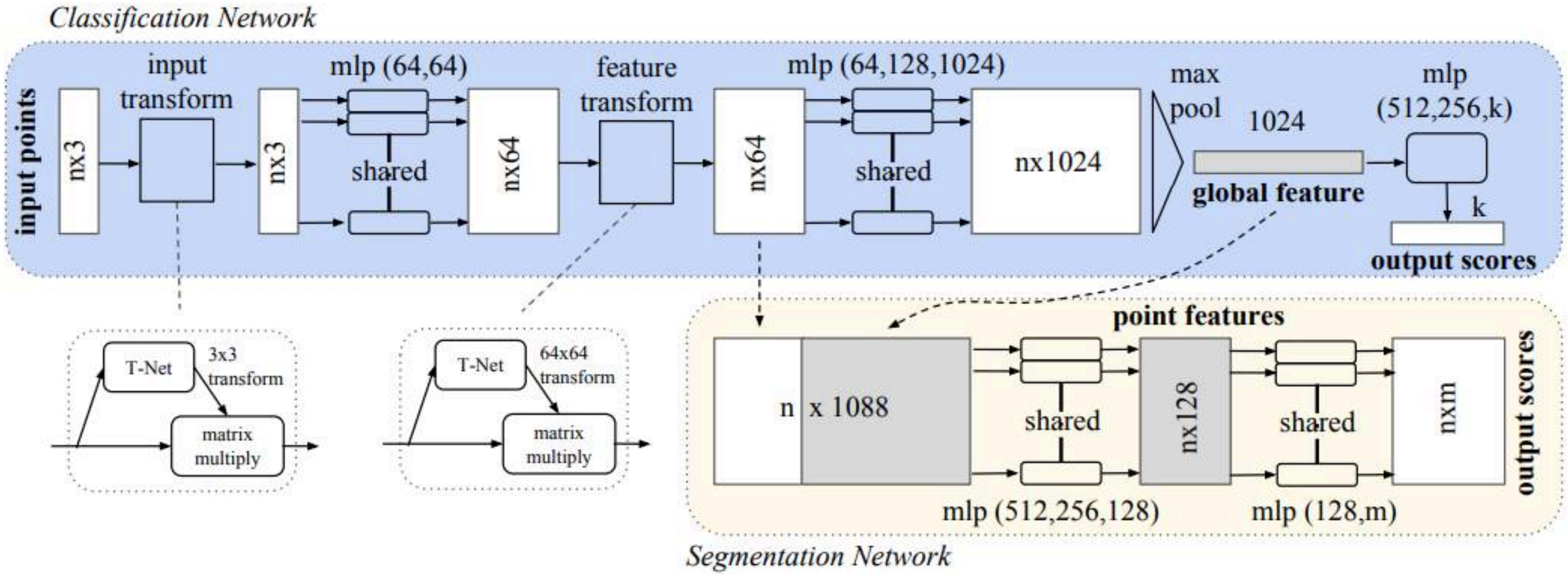
Novelty:

- Object detection is reformulated as a regression problem from image space to bounding-box coordinate space
- Single stage object detectors
  - Feature extraction, detection, classification performed in one go
- Contextual information is encoded within each prediction

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788).

Deep Learning for LIDAR data
PointNet (2017)

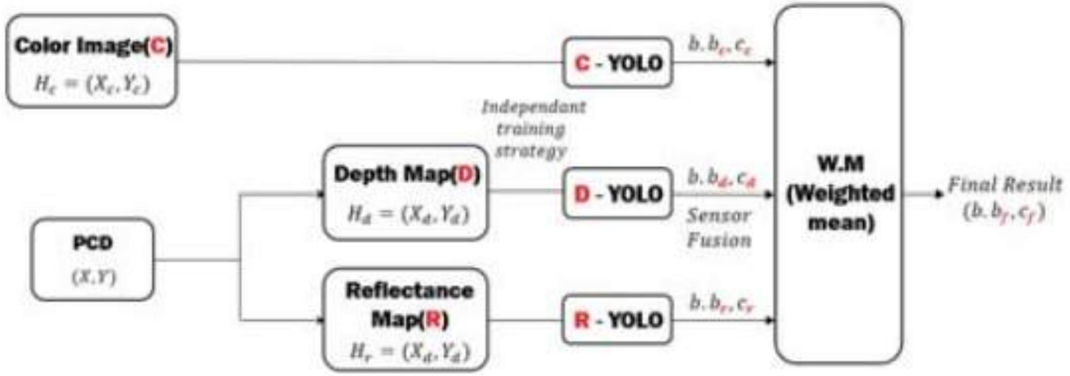
The challenge in utilizing LIDAR data is the volume of point cloud data and the permutation of their processing



- Performed classification and segmentation on *n* points of LIDAR data. Input *nx3* refers to n points with $\{x, y, z\}$ coordinate dimensions
- Used RNNs to overcome the permutation issues within LIDAR data

Qi, Charles R., et al. "Pointnet: Deep learning on point sets for 3d classification and segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017.

Deep Learning for Sensor Fusion
Vision and LIDAR



YOLO Framework is used to independently extract features from cameras and LIDAR sensors and fused to detect missed boxes

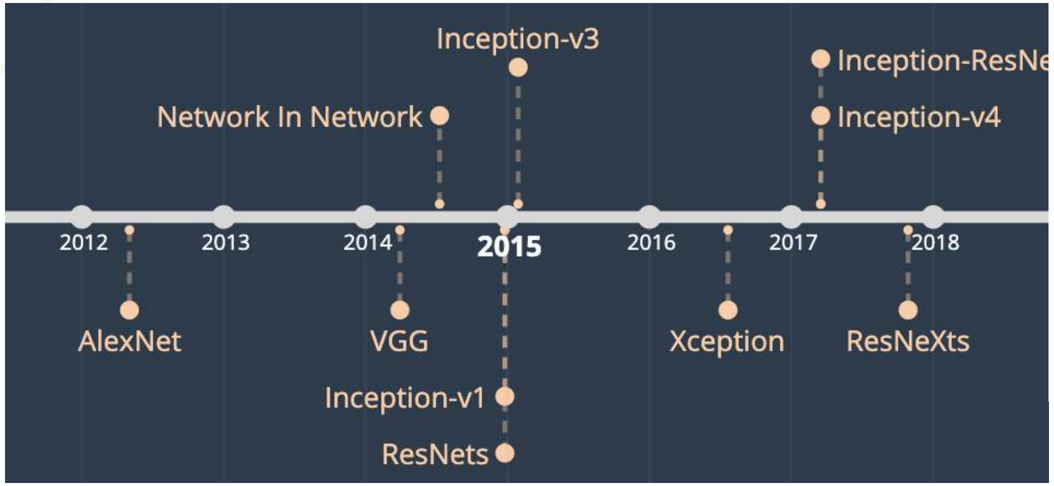
This is *'late fusion'*, in the sense that each sensor modality is independently evaluated

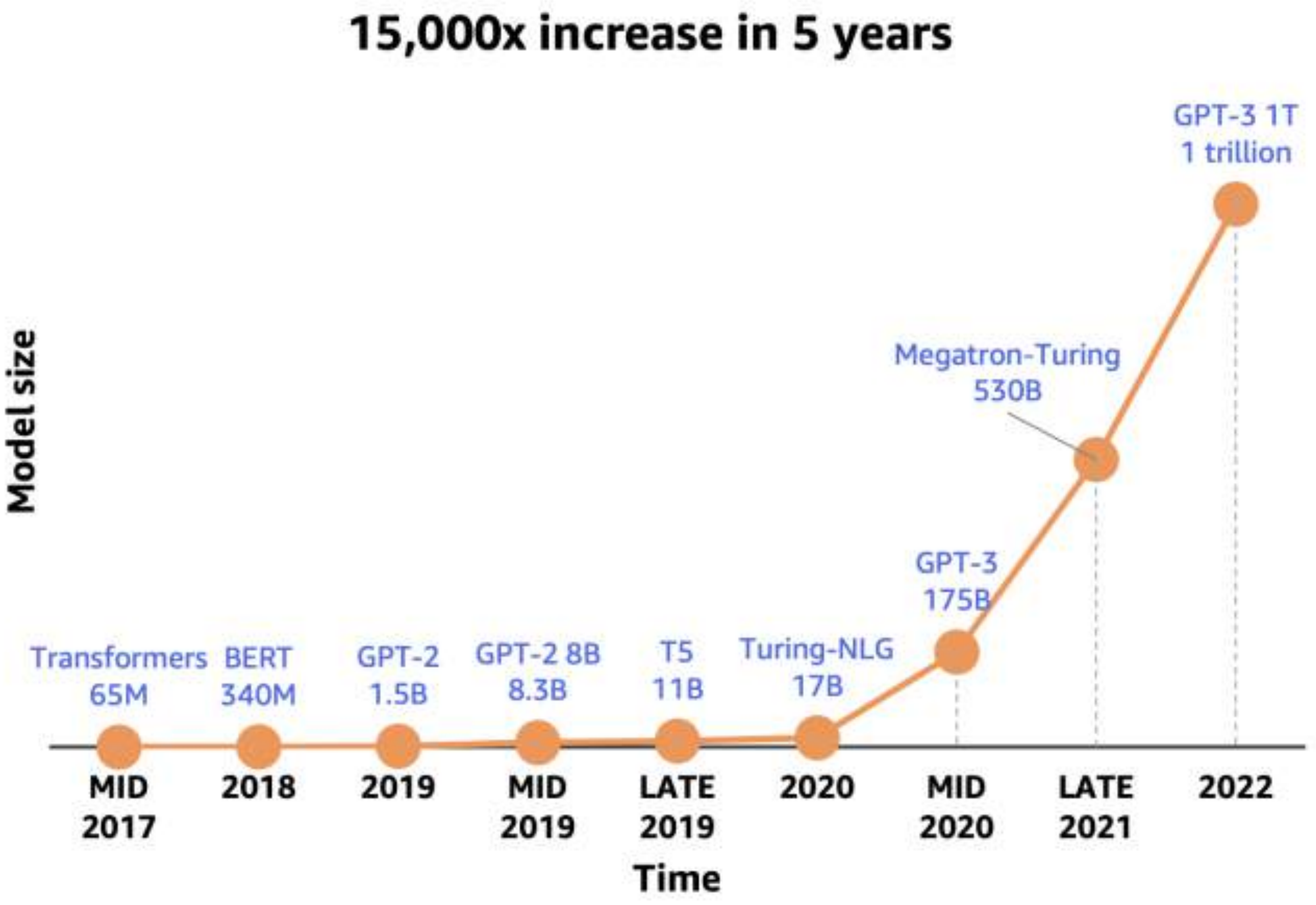
Kim, J., Kim, J., & Cho, J. (2019, December). An advanced object classification strategy using YOLO through camera and LiDAR sensor fusion. In *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)* (pp. 1-5). IEEE.

Deep Deep Deep … Deep Deep Learning

Recent Advancements

The number of parameters in models has increased exponentially





Deep Deep Deep … Deep Deep Learning

Motivation

Underlying features among different vision tasks are similar

This similarity leads to Transfer Learning

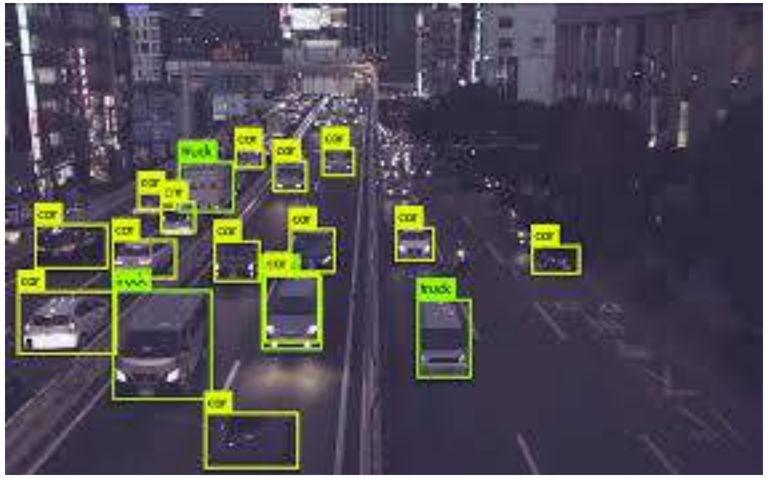
Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258* (2021).

Transfer Learning
What is Transfer Learning?

- Deep networks tend to **learn common representations** for various tasks in their earlier layers
- Can be exploited **to transfer representations from networks trained on large datasets** on one task (i.e., Image Classification on ImageNet) called the *source* to a different task called the *target* task
- Usually done by **taking large pretrained network** and then **finetuning last layer** (with all other layers frozen) on target dataset
- **Pre-trained frozen backbone** acts as a **feature extractor** while **finetuned last layer** acts to project the representations into the **decision boundary for the target task**
- Utility depends on how closely related the source and target datasets and/or tasks are

Transfer Learning
Foundation Models

Source: https://gluon-cv.mxnet.io/



Source: https://www.move-lab.com/blog/tracking-things-in-object-detection-videos

Pretraining

Foundation Model

Finetuning

Source: https://www.saagie.com/blog/object-detection-part1/

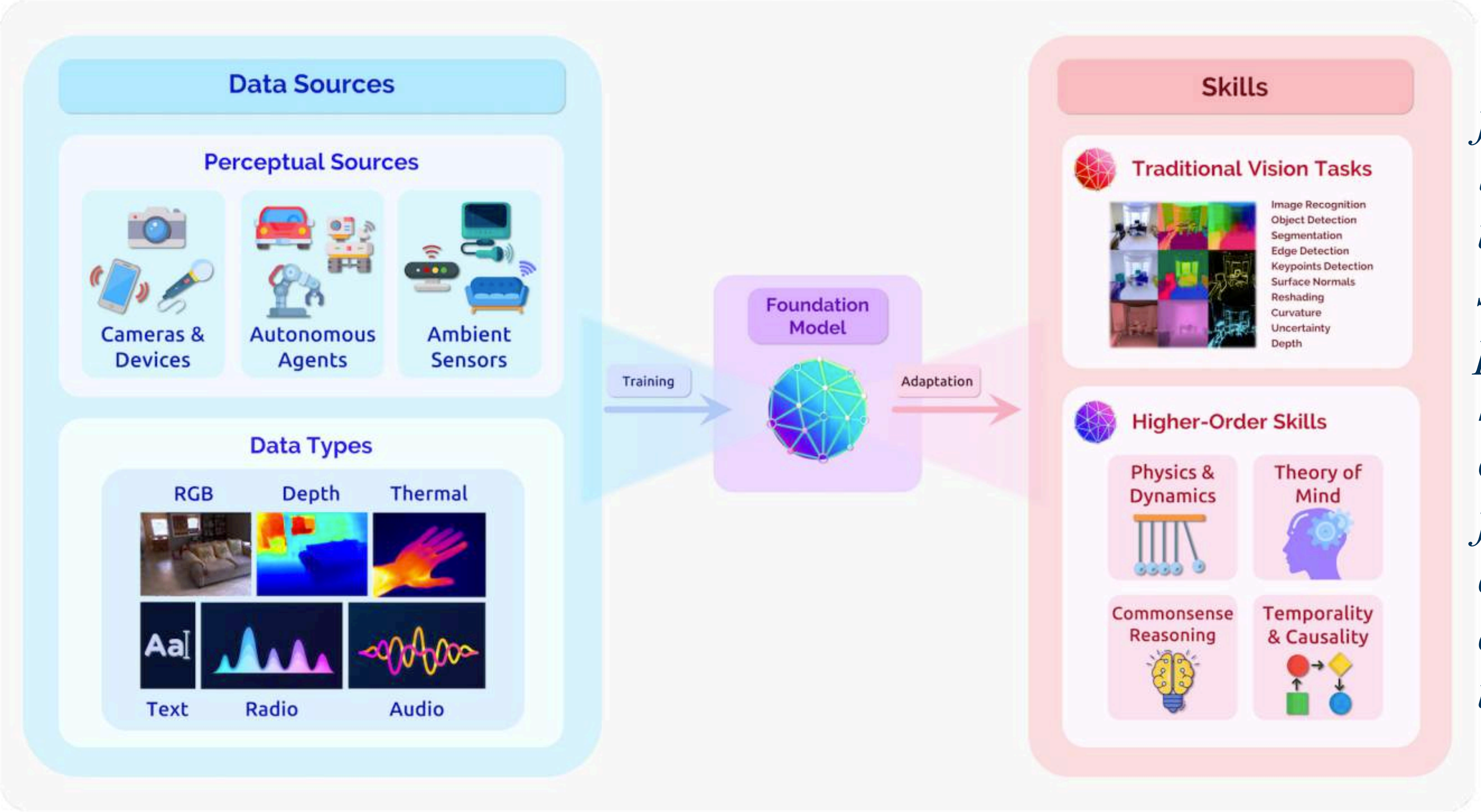
Foundation Models

Origin of the term Foundation Models

- **Foundation models** are like any other deep network that have employed **transfer learning**, except **at *scale***
- ***Scale*** brings about ***emergent* properties** that are common between tasks
- **Before 2019:** Base architectures that powered multiple neural networks were **ResNets, VGG** etc.
- **Since 2019: BERT, DALL-E, GPT, Flamingo**
- Changes since 2019: **Transformer architectures and Self-Supervision**

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258* (2021).

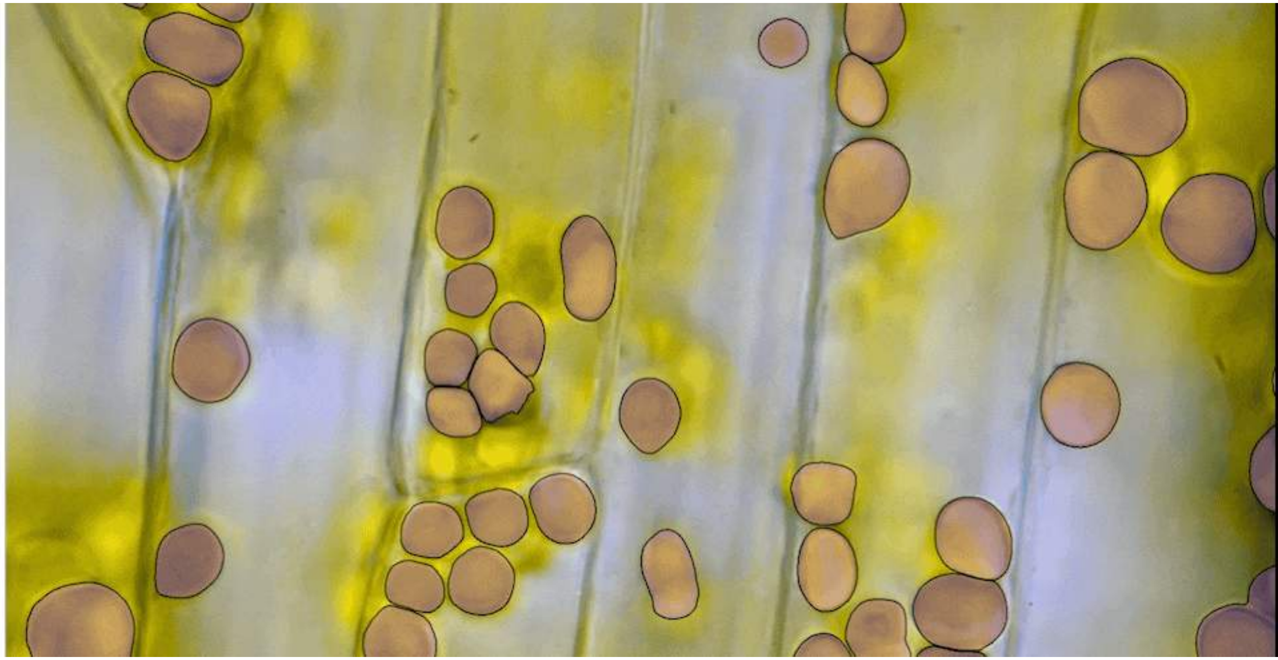
Foundation Models
Origin of the term Foundation Models



*'By harnessing **self-supervision at scale**, foundation models for vision have the potential to **distill raw, multimodal sensory information into visual knowledge**, which may effectively support traditional **perception tasks** and possibly enable new progress on challenging higher-order skills like **temporal and commonsense reasoning** These inputs can come from a **diverse range of data sources** and application domains, suggesting promise for applications in **healthcare and embodied, interactive perception settings**'*

Bommasani, Rishi, Drew A. Hudson, Ehsan Adeli, Russ Altman, Simran Arora, Sydney von Arx, Michael S. Bernstein et al. "On the opportunities and risks of foundation models." *arXiv preprint arXiv:2108.07258* (2021).

Foundation Models
Segment Anything Model



Segment Anything Model (SAM) released by Meta on April 5, 2023 was trained on Segment Anything 1 Billion dataset with 1.1 billion high-quality segmentation masks from 11 million images

Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).

Foundation Models
Segment Anything Model

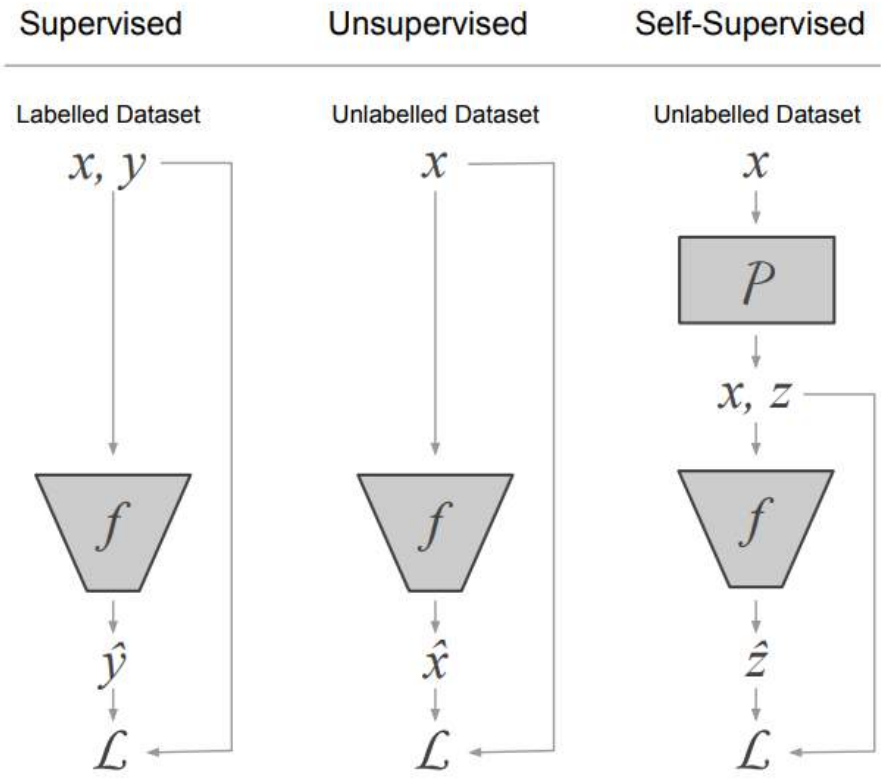
Cityscapes dataset semantic segmentation annotation took ~90 mins per image

Kirillov, Alexander, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao et al. "Segment anything." *arXiv preprint arXiv:2304.02643* (2023).

Foundation models are trained via Self-Supervision

Self-Supervision:

- Type of unsupervised learning
- Primary difference is the introduction of a **"pre-text task."**
- The pre-text task generates pseudo-labels that are used to train a network.



Ericsson, L., Gouk, H., Loy, C. C., & Hospedales, T. M. (2021). Self-Supervised Representation Learning: Introduction, Advances and Challenges. *arXiv preprint arXiv:2110.09327*.

Self-Supervision
Overall Training Process

1. **Identify Labeled and Unlabeled Data**

Unlabeled Data
$(x_1 \dots x_N)$

Labeled Data
$(x_1 \dots x_M), (y_1 \dots y_M)$

2. **Generate pseudo-labels with some pre-text task P**

Unlabeled Data
$(x_1 \dots x_N)$

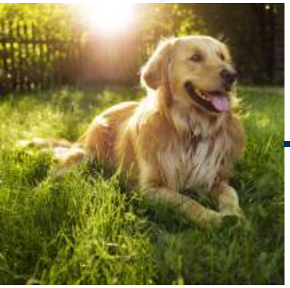
P

Pseudo - Labeled Data
$(x_1 \dots x_N), (z_1, \dots, z_M)$

Self-Supervision

Example Training Process

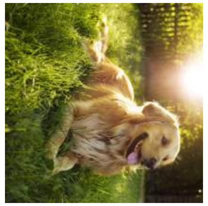
Step 1: Generate pseudo-labels via image rotations



Unlabeled Image x

- **Rotate 90** → $x_1, z_1 = 90$
- **Rotate 180** → $x_2, z_2 = 180$
- **Rotate 270** → $x_3, z_3 = 270$

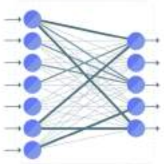
Step 2: Network learns to predict angle image is rotated



ConvNet → $\hat{z}$ → $L(\hat{z}, z_1)$ → **Update**

Step 3: Attach linear layer and train to classify labels (y) on labeled dataset

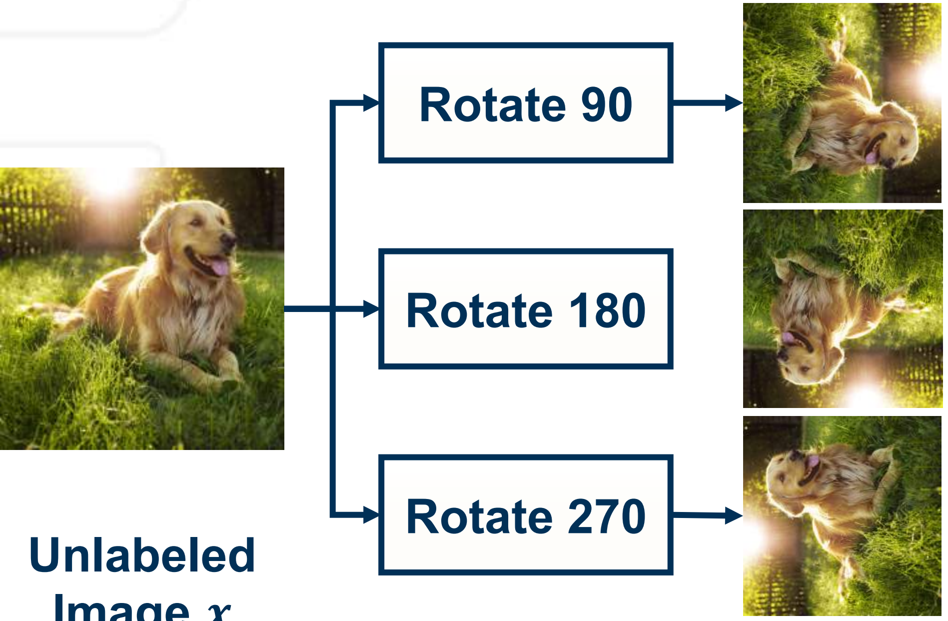
ConvNet Trained



→ $L(\hat{y}, y_1)$ → **Update**

Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*.

Step 1: Generate pseudo-labels via image rotations



Unlabeled Image x

$x_1, z_1 = 90$

$x_2, z_2 = 180$

$x_3, z_3 = 270$

Learning pre-text task will allow network to learn relevant features without needing explicit labels!

Gidaris, S., Singh, P., & Komodakis, N. (2018). Unsupervised representation learning by predicting image rotations. *arXiv preprint arXiv:1803.07728*.

Self-Supervision

Types of Pre-text Tasks

Differences in self-supervision are based on the type of pre-text task that is defined

Transformation Prediction

- Pre-text task performs some transformation on data and tasks model with trying to learn nature of transformation.

Masked Prediction

- Pre-text task removes some part of the data and the model is tasked with trying to predict what was removed.

Deep Clustering

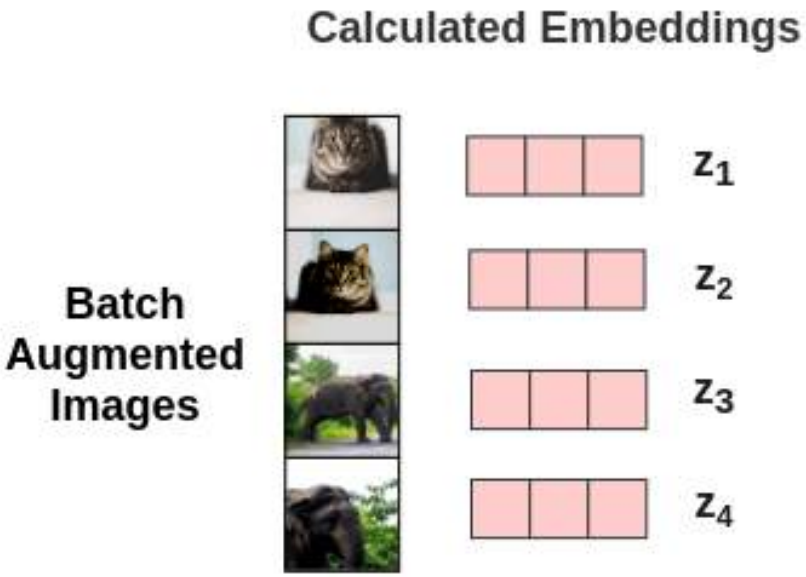
- Identify clusters of features and iteratively assign pseudo-labels to train model.

Contrastive Learning

- Pre-text task identifies positive and negative pairs of data and the model is tasked with learning similarities to discriminate between positive and negatives.

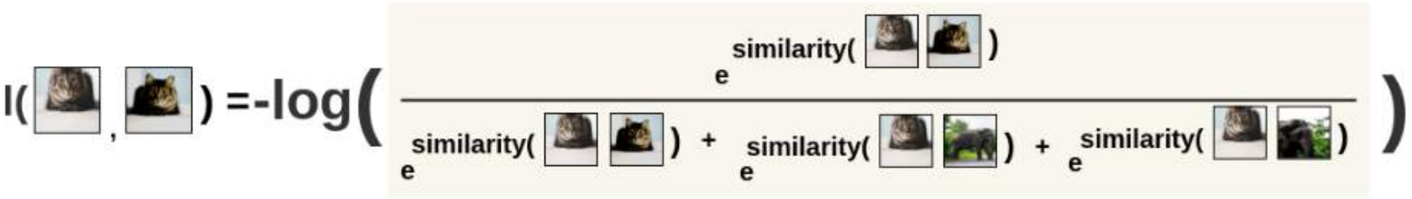
Contrastive Learning
Sim-CLR Framework

The Pseudo-labels are used to create positive-negative pairs within each batch



Contrastive loss on embeddings

$$l(i,j) = -log\frac{exp(s_{i,j})}{\sum_{k=1}^{2N} l_{[k!=i]} exp(s_{i,k})}$$

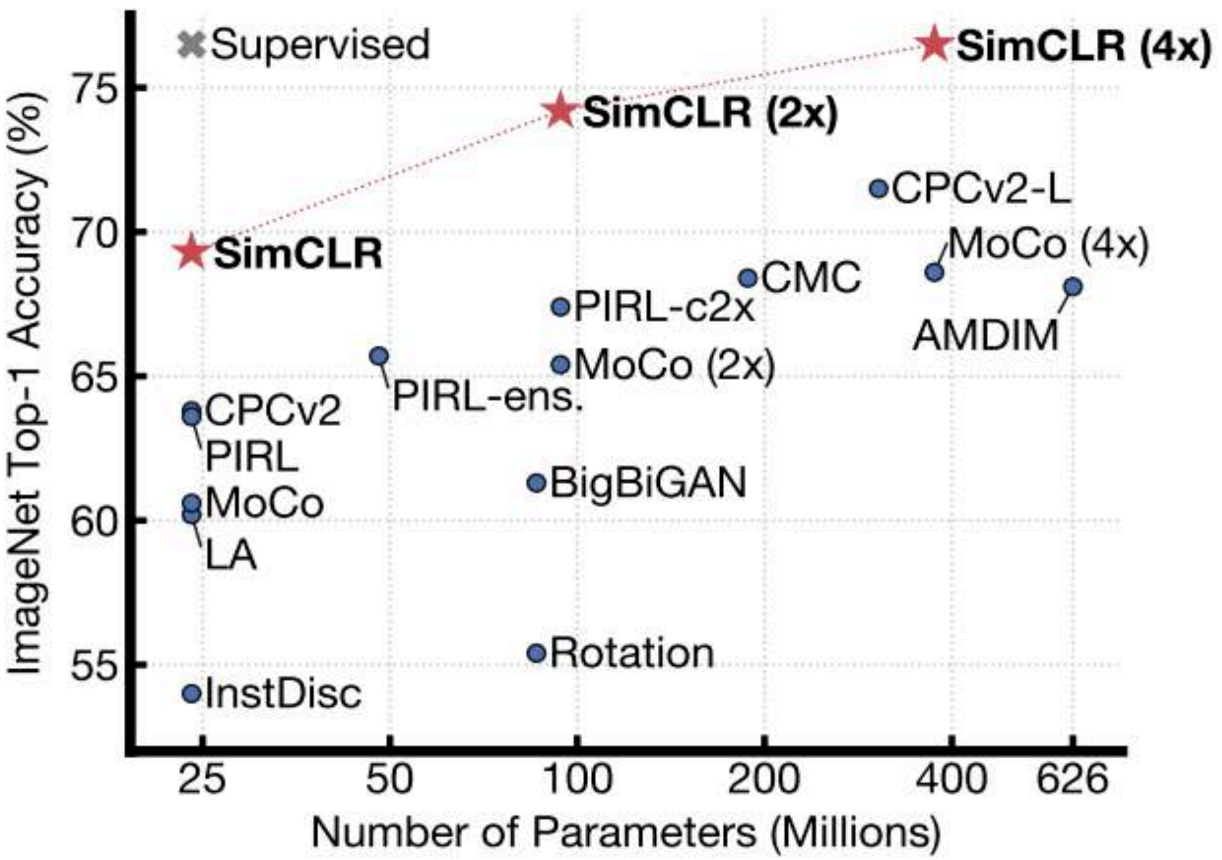


Note: The positive pairs are only the augmentations and negative pairs are all other images in the batch

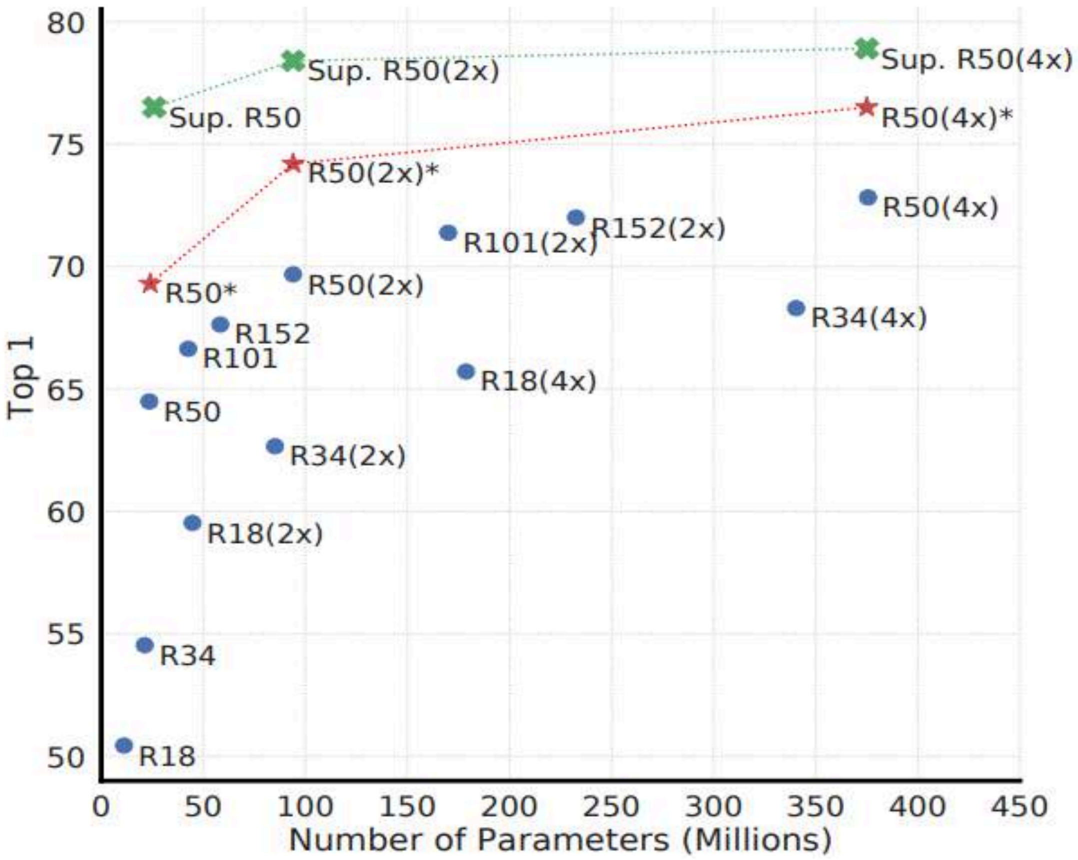
Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *arXiv preprint arXiv:2002.05709* (2020).

Contrastive Learning
Contrastive Learning vs Supervised Learning

Performance vs Models



Performance vs Parameters



Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *arXiv preprint arXiv:2002.05709* (2020).

Contrastive Learning

Contrastive Learning other than SIM-CLR

What differentiates other Contrastive Learning methods from Sim-CLR?

The way that similar pairs (positives) and dissimilar pairs (negatives) are generated.

Paper	Short description	Topics of contribution
Becker and Hinton [8]	Maximise MI between two views	Foundation
Bromley et al. [11]	Siamese network in metric learning setting	Foundation
Chopra, Hadsell, and LeCun [20]	Learn similarity metric with contrastive pair loss	Energy-based loss, Application
Hadsell, Chopra, and LeCun [39]	Learn invariant representation from pair loss	Energy-based loss, Application
Weinberger, Blitzer, and Saul [108]	Learn distance metric with triplet loss	Energy-based loss
Collobert and Weston [21]	Learn language model with triplet loss	Application
Chechik et al. [15]	Learn image retrieval model with triplet loss	Application
Noise Contrastive Estimation [38]	Introduce NCE, a general methods to learn unnormalised probabilistic model	Probabilistic loss
Mnih and Teh [71]	Learn language model with NCE-based loss	Application
Mikolov et al. [68]	Learn word embedding with Negative Sampling (NEG), a modified version of NCE	Probabilistic loss, Application
Wang et al. [105]	Learn fine-grained image similarity using deep network and triplet loss	Application
Wang and Gupta [107]	Use video's sequential coherence to learn unsupervised video representation	Similarity, Application
Lifted-structure loss [75]	Extend triplet loss to multiple positive and negative pairs per query	Energy-based loss
N-pair loss [92]	Proposed non-parametric classification loss with multiple negative pairs per query	Probabilistic loss
Wu et al. [109]	Focus on the quality of negative samples through a distance-weighted margin loss	Similarity, Energy-based loss
Hermans, Beyer, and Leibe [45]	State the important of mining hard samples in triplet loss	Similarity
Wu et al. [110]	Self-supervised representation with instance discrimination Memory bank to holds keys for next epoch	Application Encoder
CPC [77]	Mutual Information with the contrastive loss Define similarity with past-future context-instance relationship	Mutual Information loss Similarity
DIM [46]	Evaluate multiple mutual information bound for the contrastive loss Global-local context-instance relationship	Mutual Information Loss Similarity
MoCo [43]	Use momentum encoder to store features to memory queue	Encoder
SimCLR [16]	Simplify and demonstrate large empirical improvement in instance discrimination task Focus on the use of separate heads	Application Transform heads
BYOL [34]	Learning similarity without negative samples	Loss

Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *arXiv preprint arXiv:2002.05709* (2020).

IEEE Open Journal of Signal Processing

Exploiting the Distortion-Semantic Interaction in Fisheye Data

Kiran Kokilepersaud, PhD Student

Mohit Prabhushankar, PhD Postdoc

Ghassan AlRegib, PhD Professor

SCAN ME

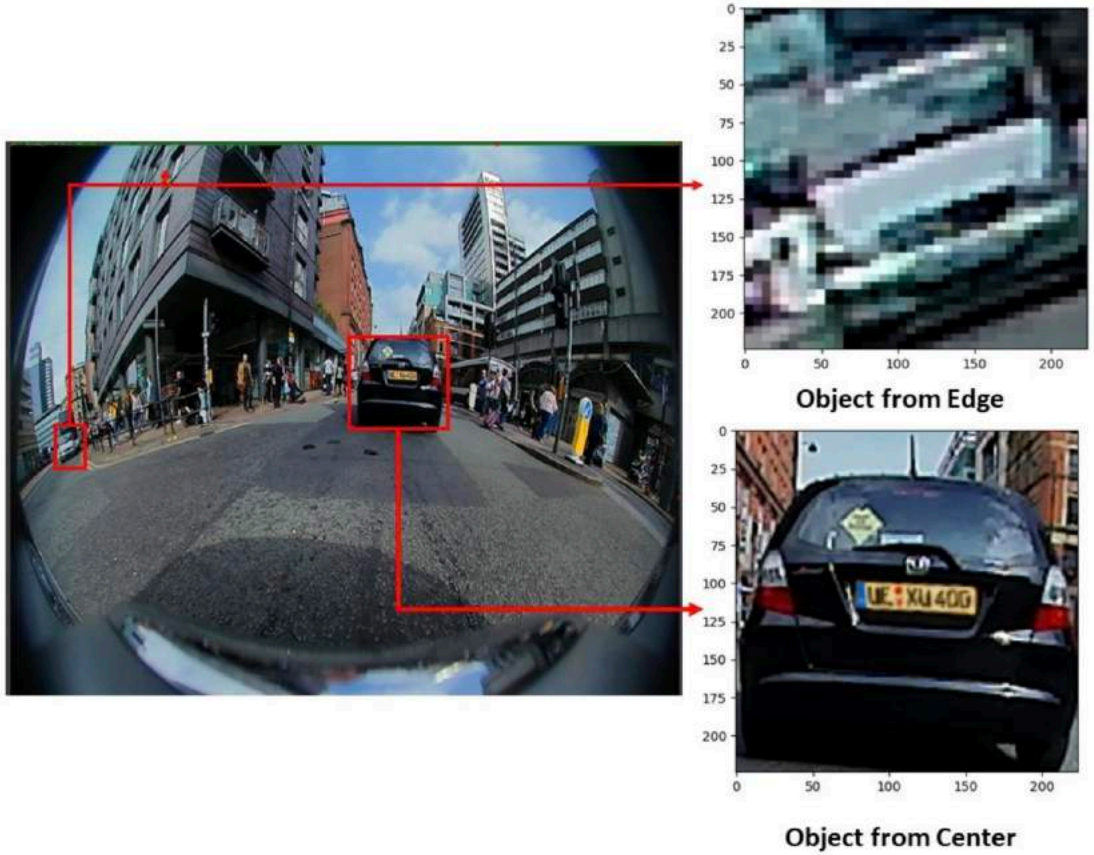
Contrastive Learning for Fisheye Images

Positive-negative pairs in Fisheye Images

Exploiting the Distortion-Semantic Interaction in Fisheye Data

SCAN ME

Intuition: Regions within a fisheye image are their own class. Hence, any object within them are positives



Intuition for Loss 1:

All objects from the edge (be it a car, bike, pedestrian) are positives and objects from the centre (be it a car, bike, pedestrian) are negatives

Intuition for Loss 1:

All objects from labeled car (be it in the center or the edge) are positives and all other objects (be it in the center or the edge) are negatives

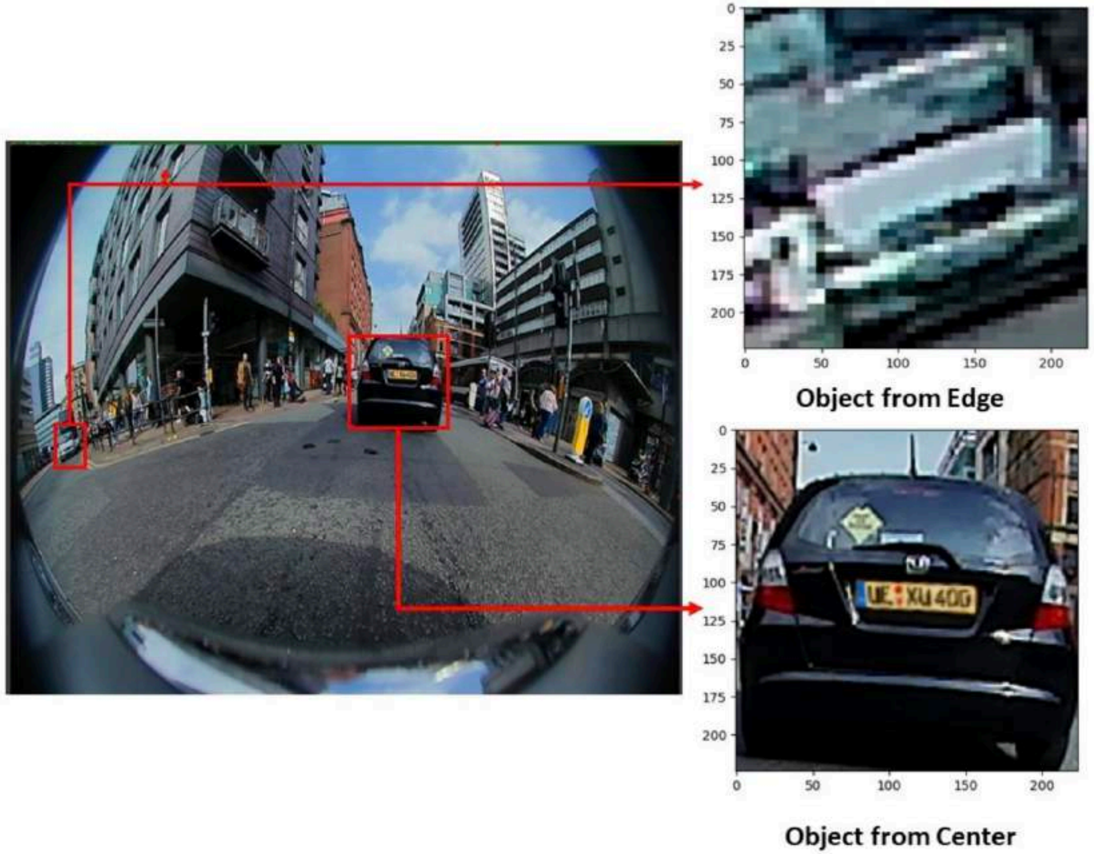
Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *arXiv preprint arXiv:2002.05709* (2020).

Contrastive Learning for Fisheye Images
Positive-negative pairs in Fisheye Images

Exploiting the Distortion-Semantic Interaction in Fisheye Data

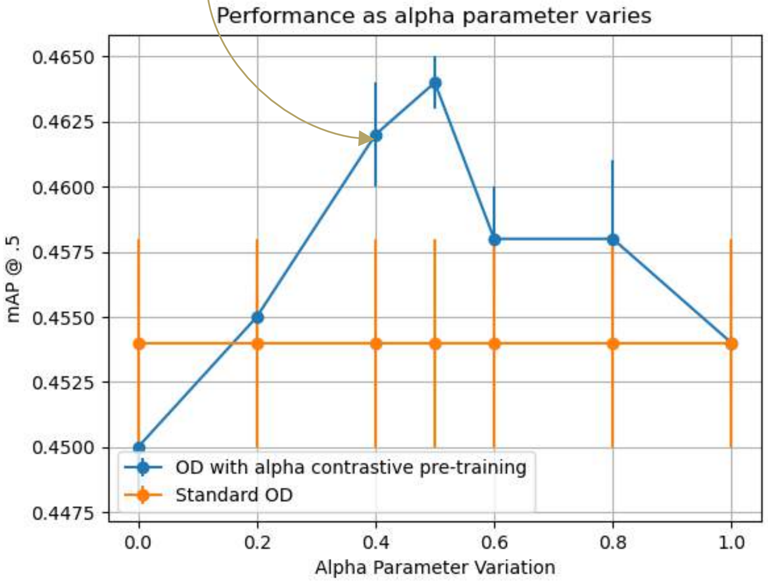
SCAN ME

Intuition: Regions within a fisheye image are their own class. Hence, any object within them are positives



$$\alpha L_{class} + (1-\alpha) L_{RegionClass}$$

α controls the level of unsupervised contrastive learning



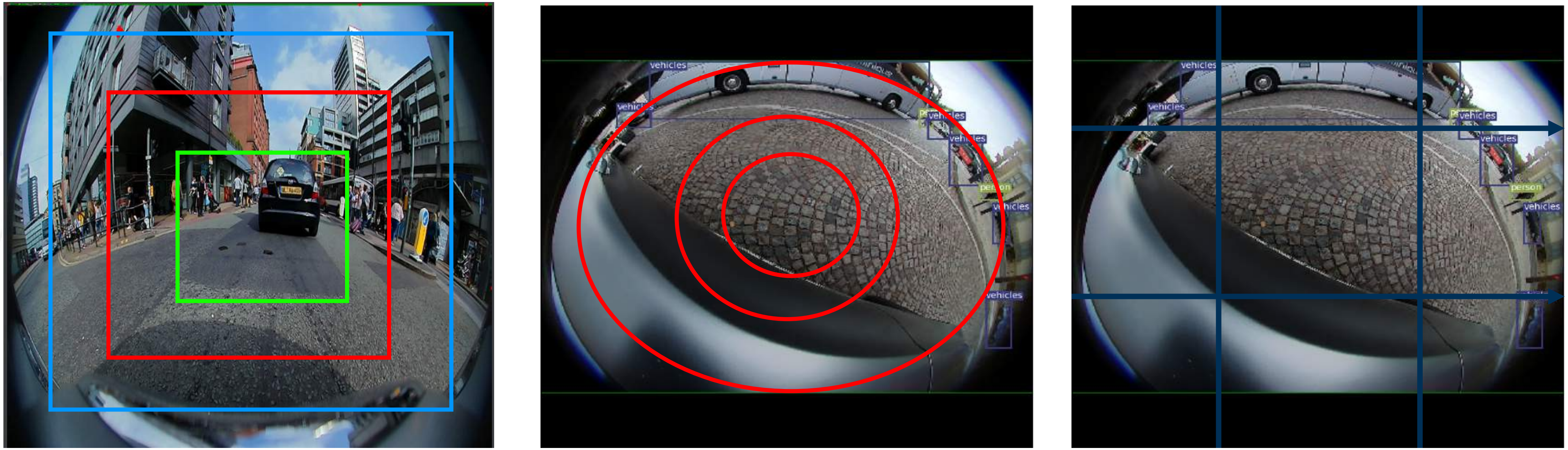
Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *arXiv preprint arXiv:2002.05709* (2020).

Contrastive Learning for Fisheye Images
Positive-negative pairs in Fisheye Images

Exploiting the Distortion-Semantic Interaction in Fisheye Data

SCAN ME

Are there alternative ways of partitioning the regions?



Defining the positive-negative pairs is application dependent

[Tutorial] | [Ghassan AlRegib and Mohit Prabhushankar] | [June 4, 2023]

Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." *arXiv preprint arXiv:2002.05709* (2020).

Objectives

Takeaways from Part II

- Part I: Challenges in Perception and Autonomy
- **Part II: Deep Learning for Perception**
    - Transfer Learning and training at scale are essential for foundation model development
    - Self-supervised Learning provides a framework for large scale learning on unannotated data
- Part III: Existing Deep Learning solutions to Challenges in Perception
- Part IV: Remaining Challenges and Future Directions

A Holistic View of Perception in Intel. Vehicles

Part III: Deep Learning at Inference

Objectives

Objectives in Part III

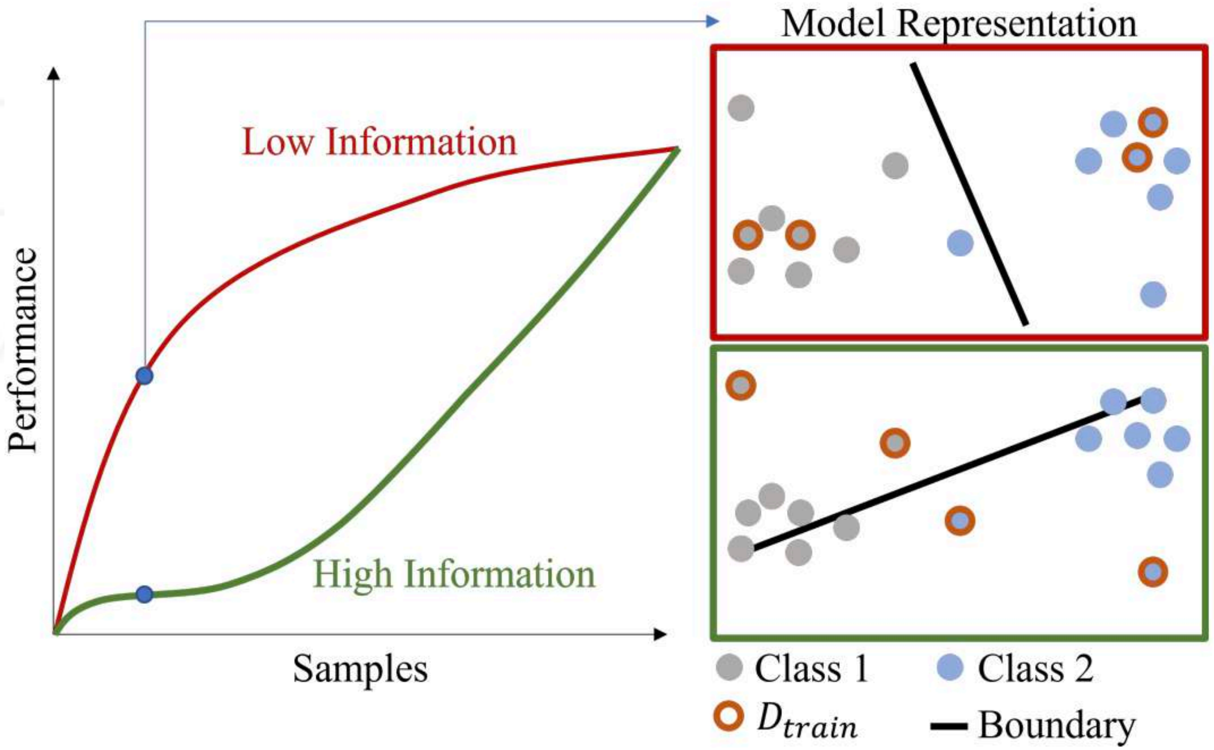
- Challenging conditions at training
- Inference
  - Deficiencies at Inference
- Overcoming deficiencies at Inference
  - Anomaly Detection
  - Uncertainty
  - Explainability
- Case study 1: Robustness to challenging conditions
- Case study 2: Aberrant Object Detection

Perception in AVs
Technical Challenges

- **Challenging weather**
- **Challenging sensing**
- **Challenging environments**
- Context awareness
- Embedded perception
- V2X perception

Challenging Conditions in Deep Learning
Integrating Challenging Conditions in Training

The most novel/aberrant samples should not be used in early training



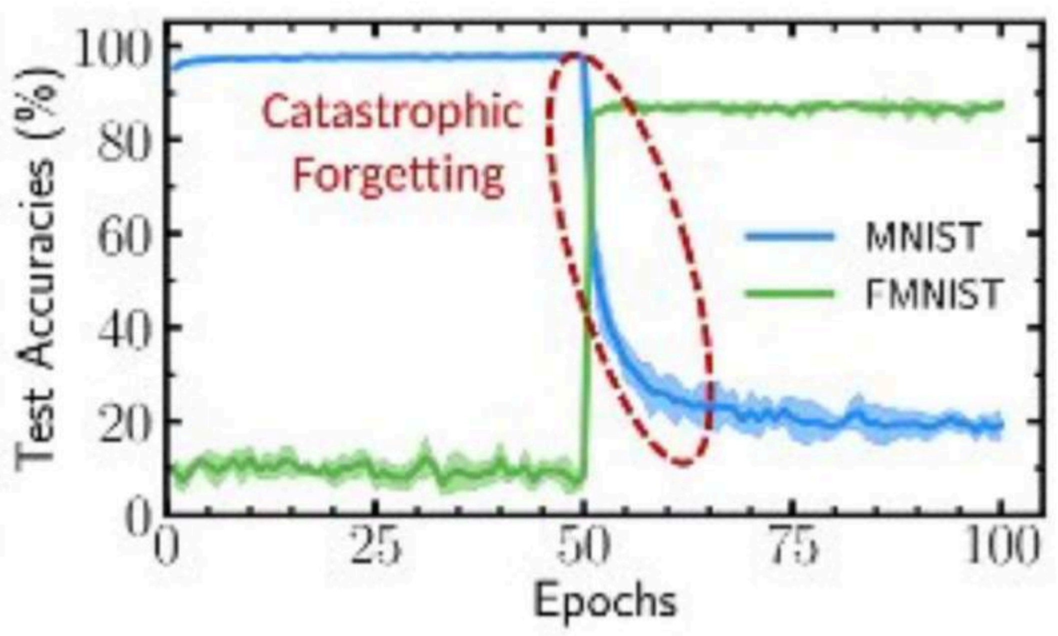
- The first instance of training must occur with less informative samples
- Less informative:
    - Highway scenarios
    - Parking
    - No accidents
    - No aberrant events

Novel samples = Most Informative

Benkert, R., Prabushankar, M., AlRegib, G., Pacharmi, A., & Corona, E. (2023). Gaussian Switch Sampling: A Second Order Approach to Active Learning. *IEEE Transactions on Artificial Intelligence*.

Integrating Challenging Conditions in Training

Subsequent training must not focus only on novel data



Catastrophic Forgetting

- The model performs well on the new scenarios, while forgetting the old scenarios
- A. number of techniques exist to overcome this trend
- However, they affect the overall performance in large-scale settings
- It is not always clear **if and when** to incorporate novel scenarios in training

Handle challenging conditions at Inference!

Laborieux, Axel, et al. "Synaptic metaplasticity in binarized neural networks." *Nature communications* 12.1 (2021): 2549.

Inference
What is Inference?

Ability of a system to predict correctly on novel data

Novel data sources:

- Test distributions
- Anomalous data
- Out-Of-Distribution data
- Adversarial data
- Corrupted data
- Noisy data
- New classes
- …

Model Train

At Deployment

Inference

What is Inference?

Ability of a system to predict correctly on novel data

Novel data sources

- Test distributions
- Anomalous data
- Out-Of-Distribution data
- Adversarial data
- Corrupted data
- Noisy data
- New classes
- …

Trained Model → Cat

Inference

Deficiencies at Inference

"The best-laid plans of sensors and networks often go awry"

- Engineers, probably

Inference

Overcoming Deficiencies at Inference

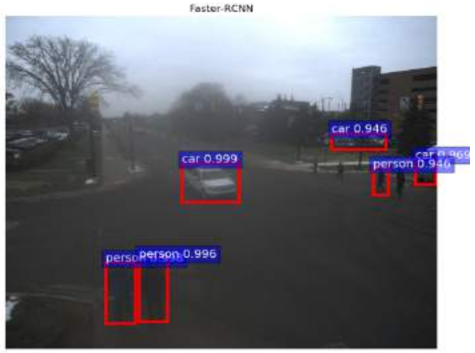
What is required when networks are met with challenging data at inference?

To overcome deficiencies, predictions from neural networks must be equipped with:

- Anomaly scores: How *close* to the training data is the novel data at inference?
- Uncertainty scores: How close to the *best* possible network is the trained network?
- Contextual Explainability: How *relevant* are the network explanations for its prediction?

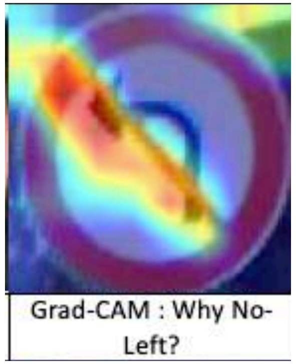
Training data

Anomalous data



Certain objects

Uncertain objects



'Why P'

'Why P, rather than Q?'

Overcoming Deficiencies at Inference

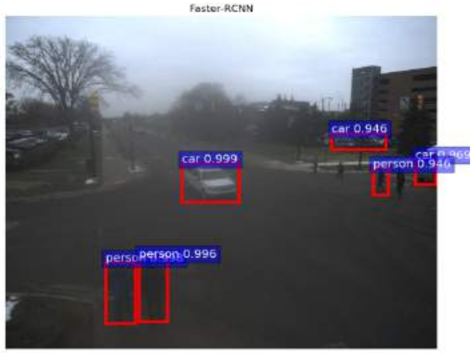
What is required when networks are met with challenging data at inference?

To overcome deficiencies, predictions from neural networks must be equipped with:

- **Anomaly scores**: How *close* to the training data is the novel data at inference?
- Uncertainty scores: How close to the *best* possible network is the trained network?
- Contextual Explainability: How *relevant* are the network explanations for its prediction?

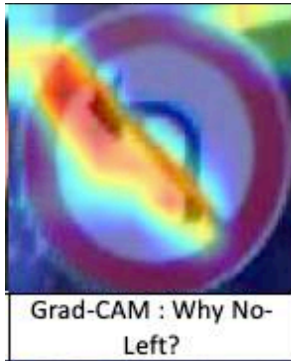
Training data

Anomalous data



Certain objects

Uncertain objects



'Why P'

'Why P, rather than Q?'

Backpropagated Gradient Representations for Anomaly Detection



Gukyeong Kwon, PhD
Amazon AWS

Mohit Prabhushankar, PhD
Postdoc, Georgia Tech

Ghassan AlRegib, PhD
Professor, Georgia Tech

SCAN ME

Anomalies

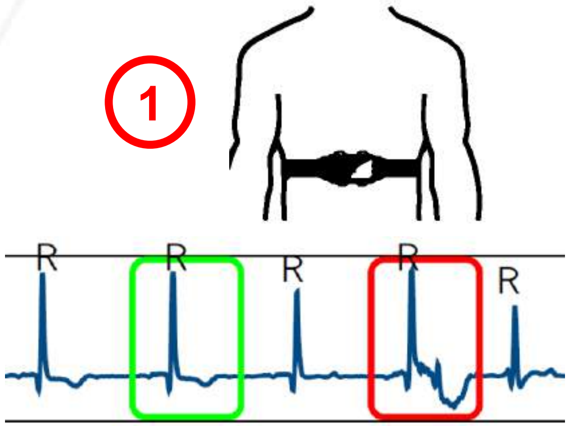
Finding Rare Events in Normal Patterns

SCAN ME

Backpropagated Gradient Representations for Anomaly Detection

'Anomalies are patterns in data that do not conform to a well defined notion of normal behavior' [1]

1



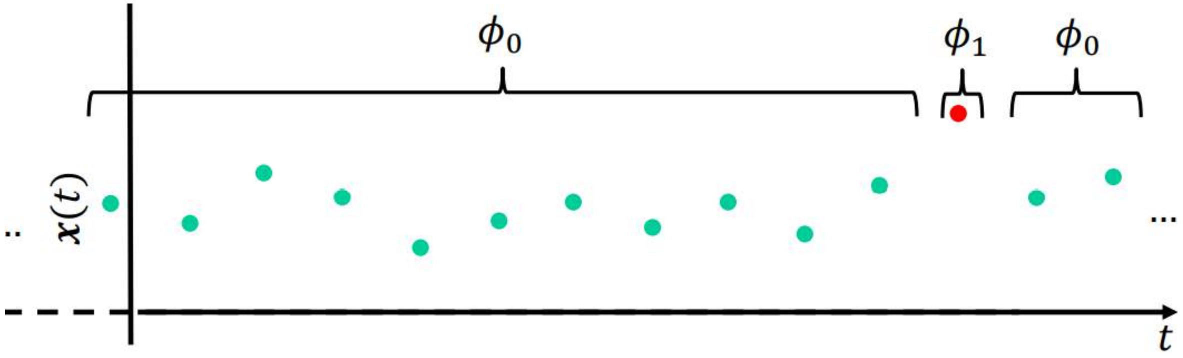
2

Statistical Definition:

- Normal data are generated from a stationary process P_N
- Anomalies are generated from a different process $P_A \neq P_N$

Goal: Detect ϕ_1

$$x(t) = \begin{cases} \phi_0 & \text{Normal data} \\ \phi_1 & \text{Anomalies} \end{cases}$$



[1] V. Chandola, A. Banerjee, V. Kumar. "Anomaly detection: A survey". ACM Comput. Surv. 41, 3, Article 15 (July 2009), 58 pages

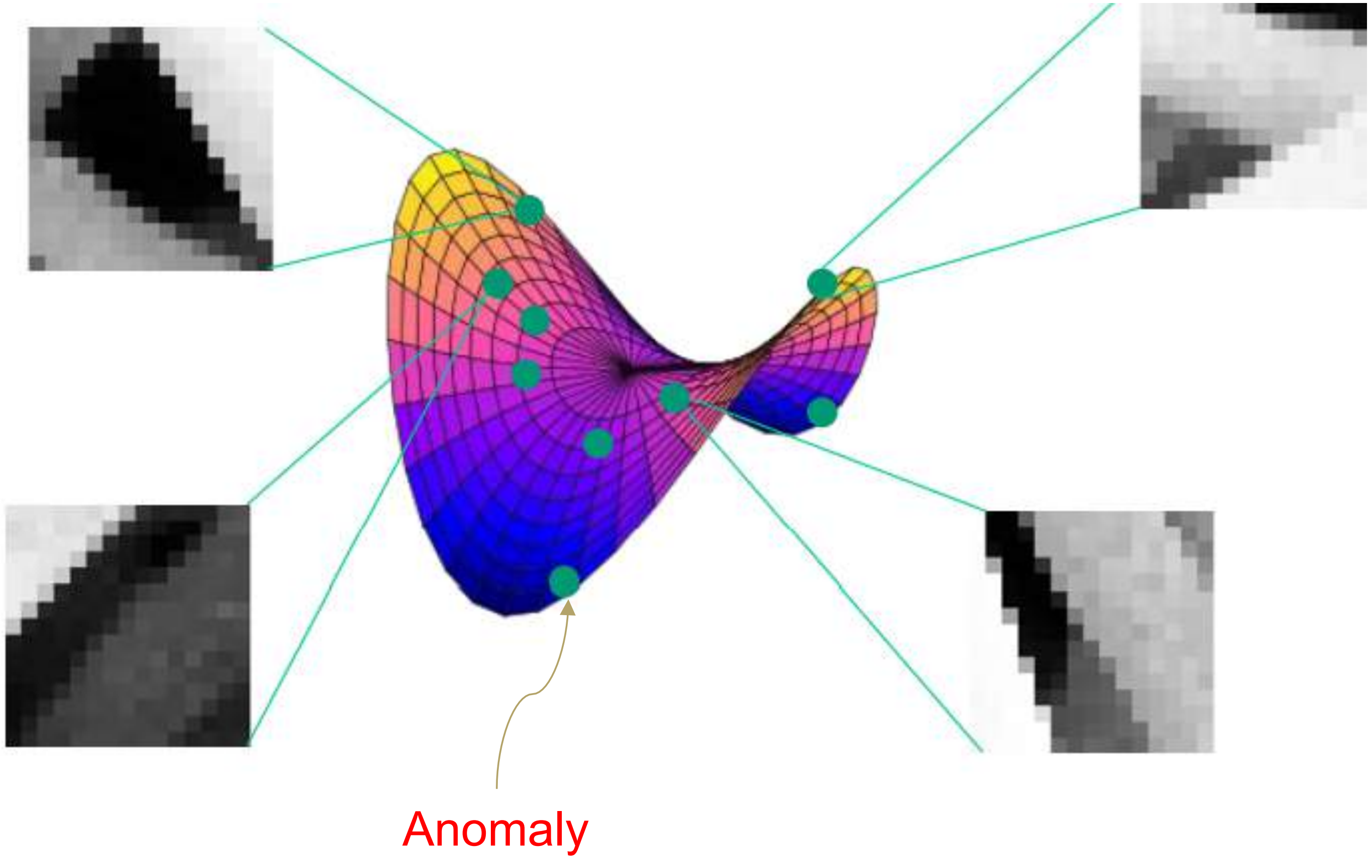
Steps for Anomaly Detection

Backpropagated Gradient Representations for Anomaly Detection

SCAN ME

Step 1: Constrain manifolds, Step 2: Detect statistically implausible projections

- Step 1 ensures that patches from natural images live close to a low dimensional manifold
- Step 2 designs distance functions that detect *implausibility* based on constraints



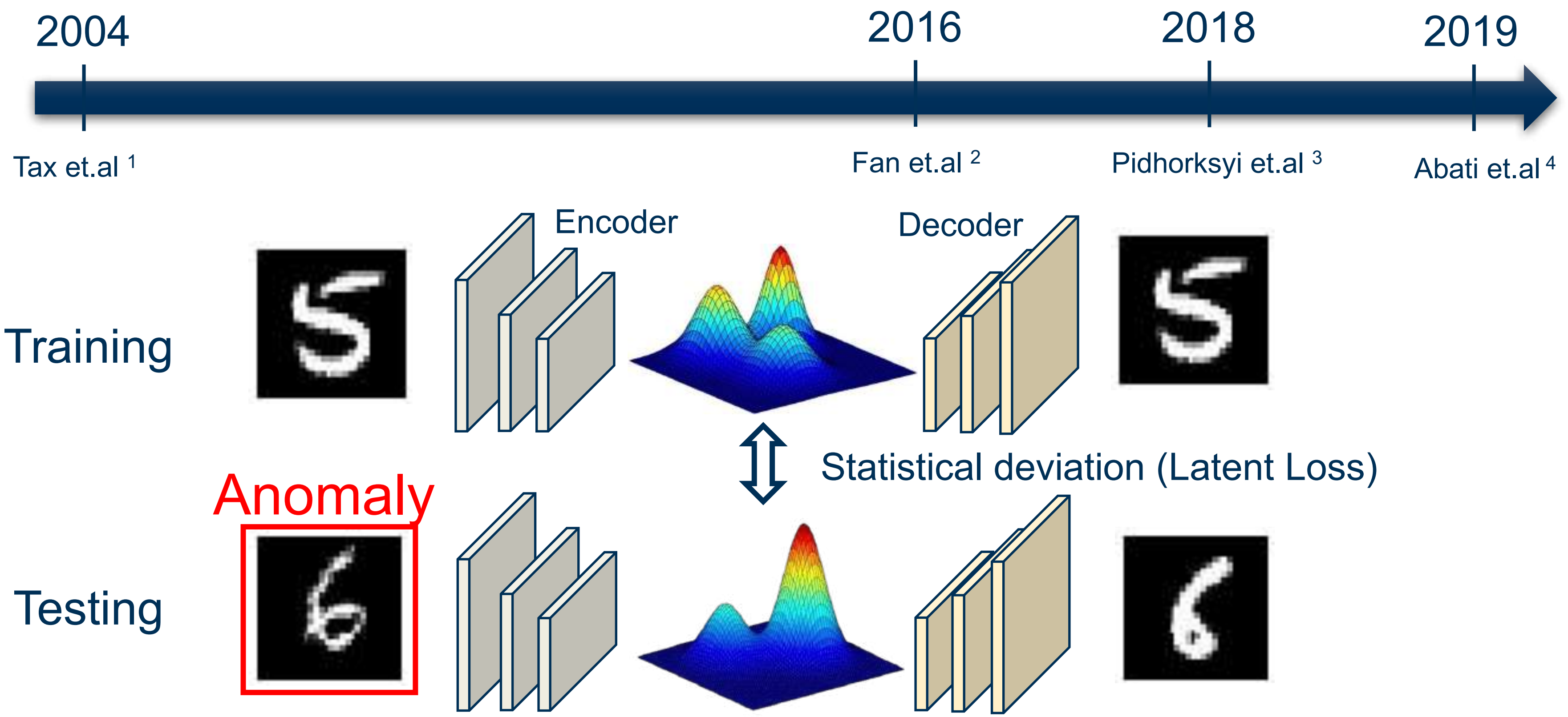
Constraining Manifolds
General Constraints

Backpropagated Gradient Representations for Anomaly Detection

SCAN ME

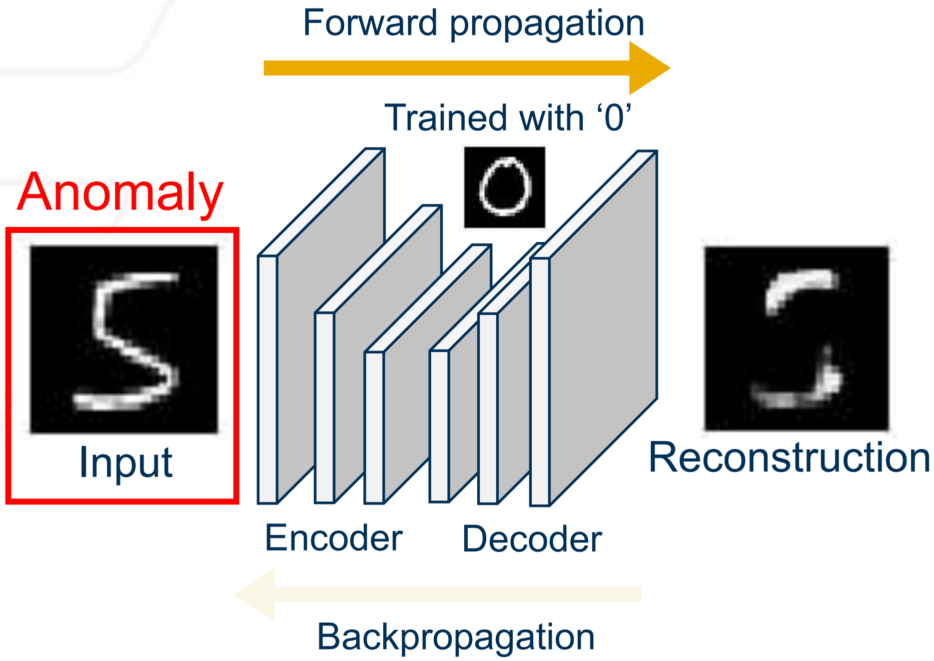
Constrained Representation

Activations are constrained using GANs, VAEs, etc.



[1] David MJ Tax and Robert PW Duin. Support vector data description. Machine learning, 54(1):45–66, 2004.
[2] Yaxiang Fan, Gongjian Wen, Deren Li, Shaohua Qiu, and Martin D Levine. Video anomaly detection and localization via gaussian mixture fully convolutional variational autoencoder. arXiv preprint arXiv:1805.11223, 2018. 1, 2
[3] S. Pidhorskyi, R. Almohsen, and G. Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," in Advances in Neural Information Processing Systems, 2018, pp. 6822–6833.
[4] D. Abati, A. Porrello, S. Calderara, and R. Cucchiara, "Latent space autoregression for novelty detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 481–490.

Constraining Manifolds
Gradient-based Constraints



Activation Constraints

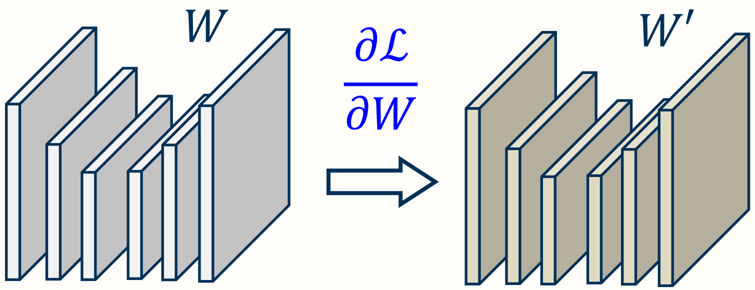
Activation-based representation
(Data perspective)

e.g. Reconstruction error ($\mathcal{L}$)

How much of the input does not correspond to the learned information?

Gradient Constraints

Gradient-based Representation
(**Model** perspective)



How much **model update** is required by the input?

G. Kwon, M. Prabhushankar, D. Temel, and G. AlRegib, "Backpropagated Gradient Representations for Anomaly Detection," 2020

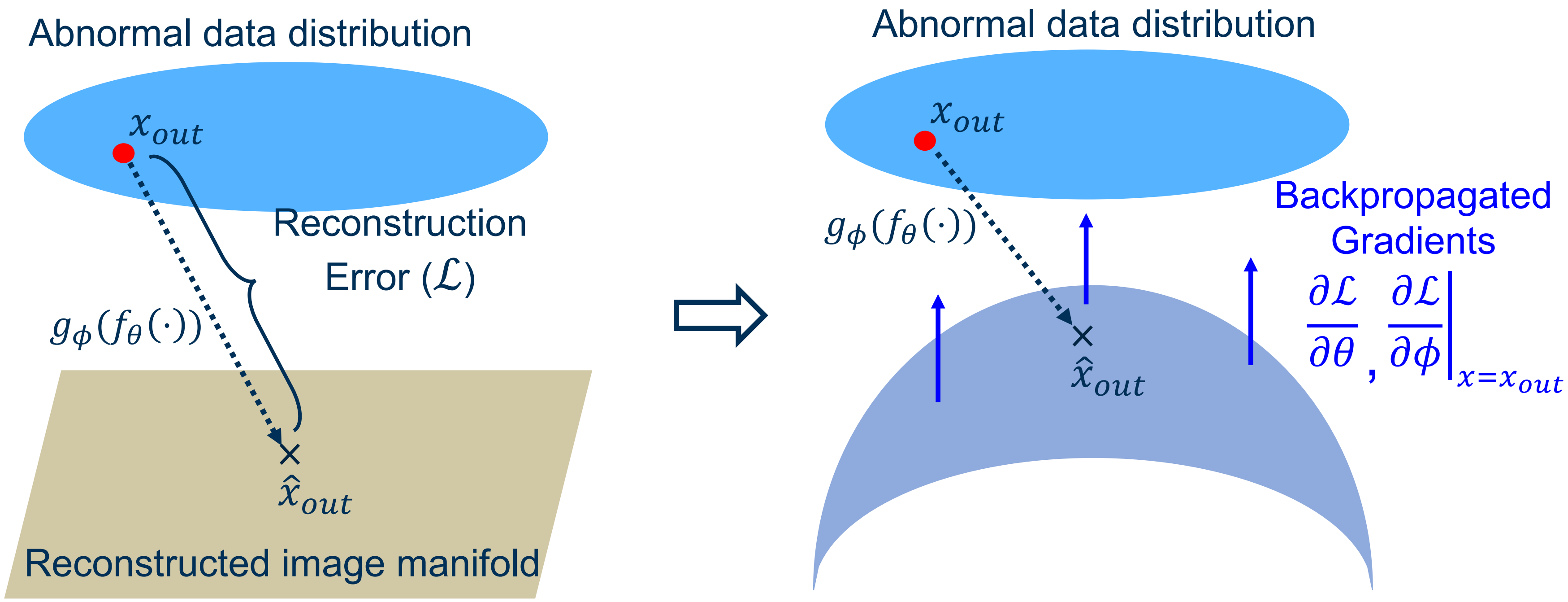
Constraining Manifolds

Advantages of Gradient-based Constraints

Backpropagated Gradient Representations for Anomaly Detection

SCAN ME

- **Gradients provide directional information to characterize anomalies**
- **Gradients from different layers capture abnormality at different levels of data abstraction**



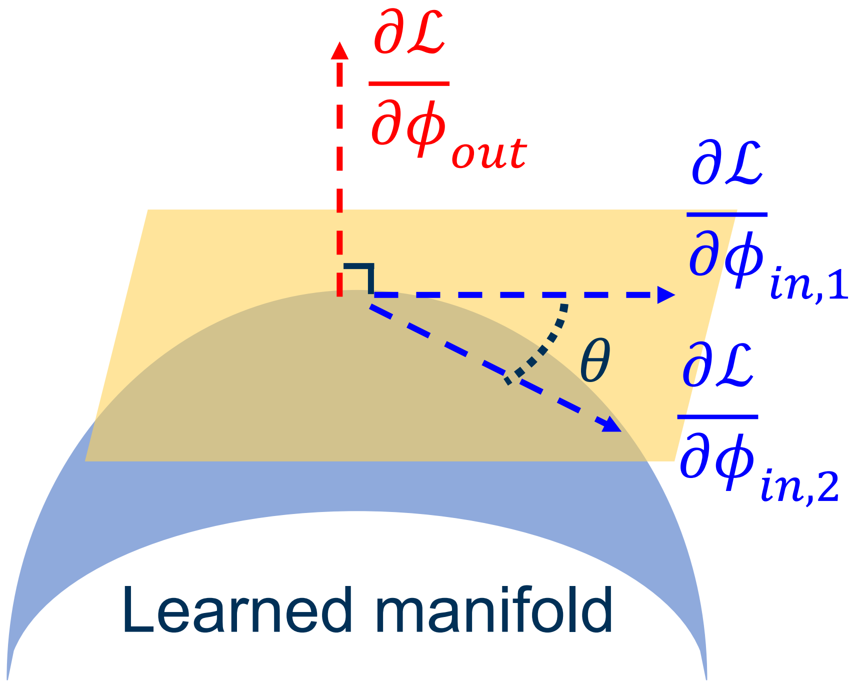
G. Kwon, M. Prabhushankar, D. Temel, and G. AlRegib, "Backpropagated Gradient Representations for Anomaly Detection," 2020

GradCON: Gradient Constraint
Gradient-based Constraints

SCAN ME

Backpropagated Gradient Representations for Anomaly Detection

Constrain gradient-based representations during training to obtain clear separation between normal data and abnormal data



ϕ: Weights $\mathcal{L}$: Reconstruction error

At k-th step of training,

$$J = \mathcal{L} - \mathbb{E}_i \left[\underbrace{\text{cosSIM}\left(\frac{\partial J}{\partial \phi_{i\,avg}}^{k-1}, \frac{\partial \mathcal{L}}{\partial \phi_i}^{k} \right)}_{\text{Gradient loss}} \right]$$

Avg. training gradients until (k-1) th iter.

Gradients at k-th iter.

where $$\frac{\partial J}{\partial \phi_{i\,avg}}^{k-1} = \sum_{t=1}^{k-1} \frac{\partial J}{\partial \phi_i}^{t}$$

G. Kwon, M. Prabhushankar, D. Temel, and G. AlRegib, "Backpropagated Gradient Representations for Anomaly Detection," 2020

GradCON: Gradient Constraint

Activations vs Gradients

Backpropagated Gradient Representations for Anomaly Detection

SCAN ME

Abnormal "class" detection (CIFAR-10)

e.g.

Normal

Abnormal

AUROC Results

Model	Loss	Plane	Car	Bird	Cat	Deer	Dog	Frog	Horse	Ship	Truck	Average
CAE	Recon	0.682	0.353	0.638	0.587	0.669	**0.613**	0.495	0.498	0.711	0.390	0.564
CAE + Grad	Recon	0.659	0.356	**0.640**	0.555	0.695	0.554	0.549	0.478	0.695	0.357	0.554
	Grad	**0.752**	0.619	0.622	0.580	0.705	0.591	0.683	**0.576**	**0.774**	**0.709**	**0.661**
VAE	Recon	0.553	0.608	0.437	0.546	0.393	0.531	0.489	0.515	0.552	0.631	0.526
	Latent	0.634	0.442	**0.640**	0.497	**0.743**	0.515	**0.745**	0.527	0.674	0.416	0.583
VAE + Grad	Recon	0.556	0.606	0.438	0.548	0.392	0.543	0.496	0.518	0.552	0.631	0.528
	Latent	0.586	0.396	0.618	0.476	0.719	0.474	0.698	0.537	0.586	0.413	0.550
	Grad	0.736	**0.625**	0.591	**0.596**	0.707	0.570	0.740	0.543	0.738	0.629	0.647

Recon: Reconstruction error, Latent: Latent loss, Grad: Gradient loss

- (CAE vs. CAE + Grad) Effectiveness of the gradient constraint
- (CAE vs. VAE) Performance sacrifice from the latent constraint
- (VAE vs. VAE + Grad) Complementary features from the gradient constraint

G. Kwon, M. Prabhushankar, D. Temel, and G. AlRegib, "Backpropagated Gradient Representations for Anomaly Detection," 2020

GradCON: Gradient Constraint
Aberrant Condition Detection

Backpropagated Gradient Representations for Anomaly Detection

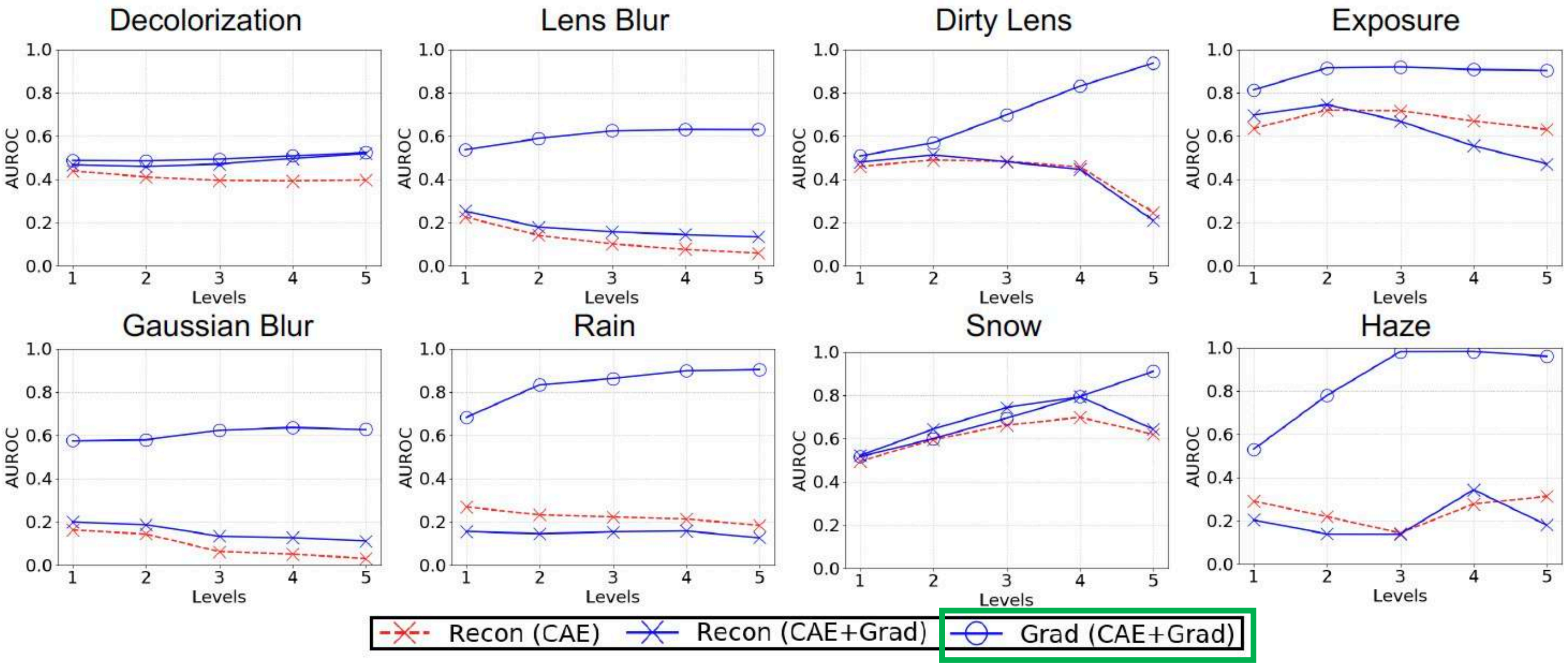
SCAN ME

Abnormal “condition” detection (CURE-TSR)

Normal

Abnormal

AUROC Results



Recon: Reconstruction error, Grad: Gradient loss

G. Kwon, M. Prabhushankar, D. Temel, and G. AlRegib, "Backpropagated Gradient Representations for Anomaly Detection," 2020

Overcoming Deficiencies at Inference

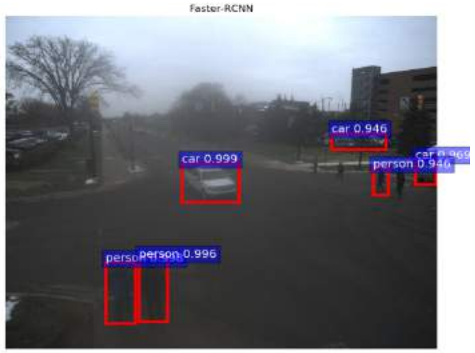
What is required when networks are met with challenging data at inference?

To overcome deficiencies, predictions from neural networks must be equipped with:

- Anomaly scores: How *close* to the training data is the novel data at inference?
- **Uncertainty scores**: How close to the *best* possible network is the trained network?
- Contextual Explainability: How *relevant* are the network explanations for its prediction?

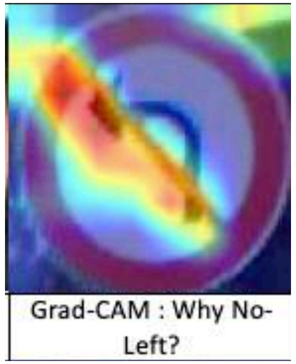
Training data

Anomalous data



Certain objects

Uncertain objects



'Why P'

'Why P, rather than Q?'

IEEE Access

Probing the Purview of Neural Networks via Gradient Analysis

Jinsol Lee,
PhD Candidate

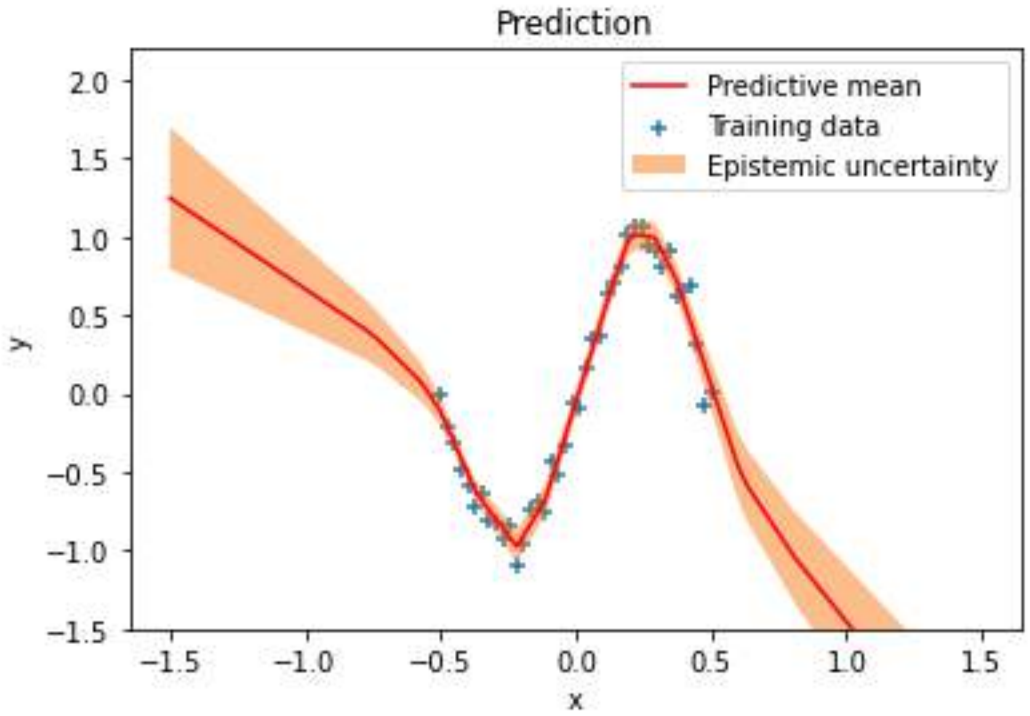
Mohit Prabhushankar, PhD
Postdoc

Ghassan AlRegib, PhD
Professor

SCAN ME

Uncertainty
What is Uncertainty?

Uncertainty is a model knowing that it does not know

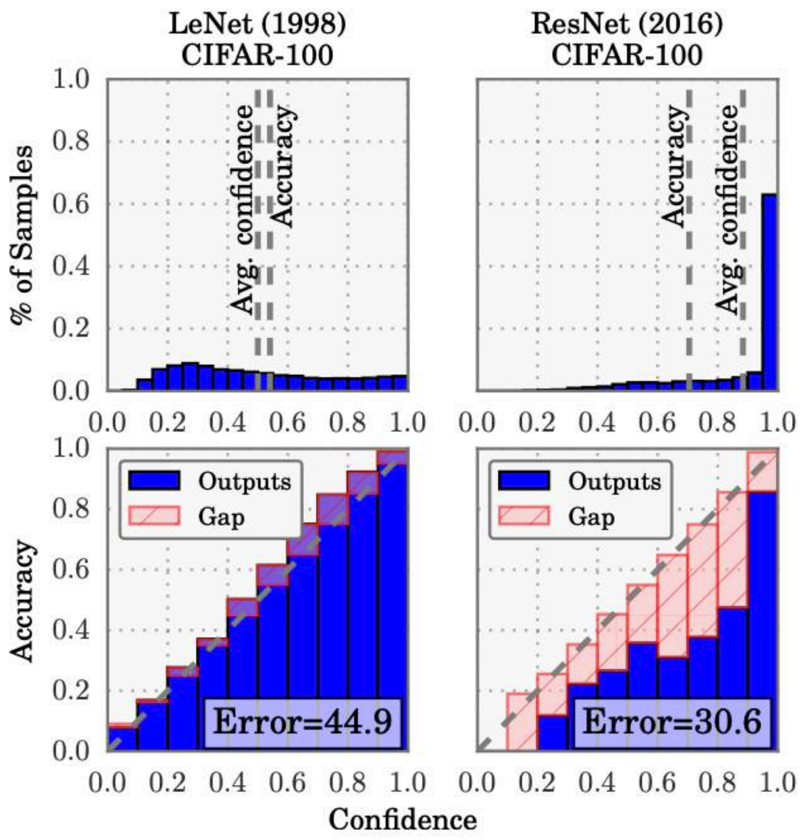


A simple example: More the training data, lesser the uncertainty

http://krasserm.github.io/2020/09/25/reliable-uncertainty-estimates/

SCAN ME

Uncertainty is a model knowing that it does not know



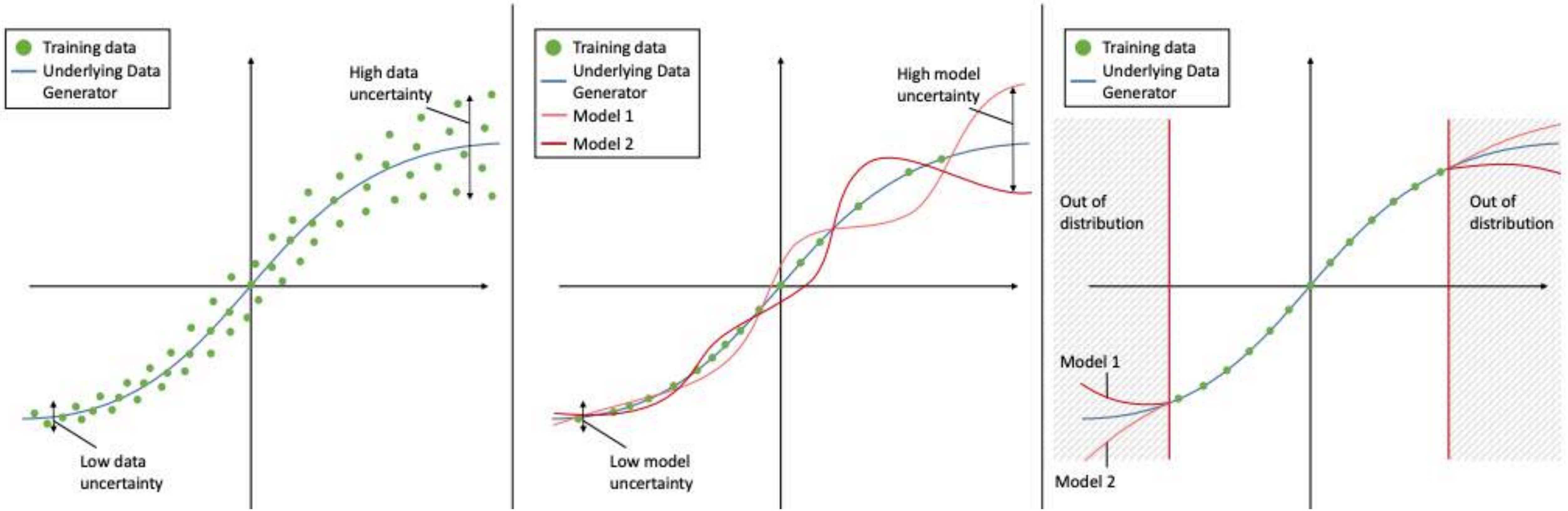
- Larger the model, more misplaced is a network's confidence
- On ResNet, the gap between prediction accuracy and its corresponding confidence is significantly high
- On OOD data, uncertainty is not easy to quantify

Guo, Chuan, et al. "On calibration of modern neural networks." *International conference on machine learning*. PMLR, 2017.

SCAN ME

Probing the Purview of Neural Networks via Gradient Analysis

Two major types of uncertainty: Uncertainty in data and uncertainty in model



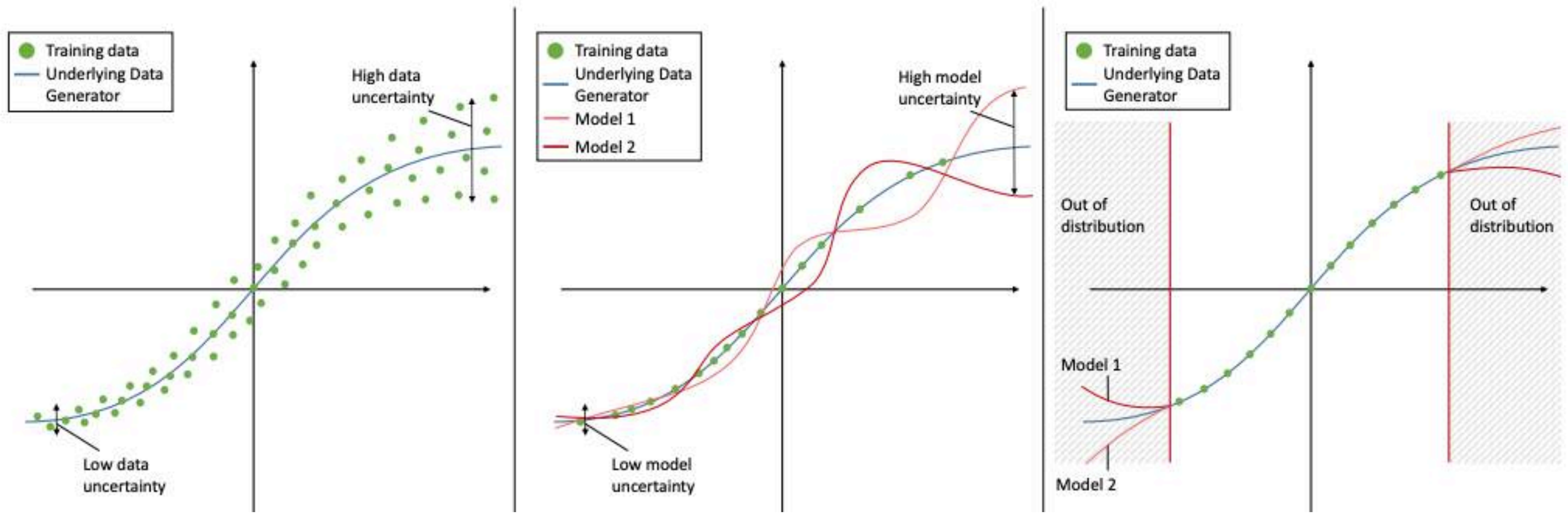
Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., ... & Zhu, X. X. (2021). A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*.

Uncertainty
Types of Uncertainty

Probing the Purview of Neural Networks via Gradient Analysis

SCAN ME

For the purpose of predictions: Both uncertainties are combined as Predictive Uncertainty

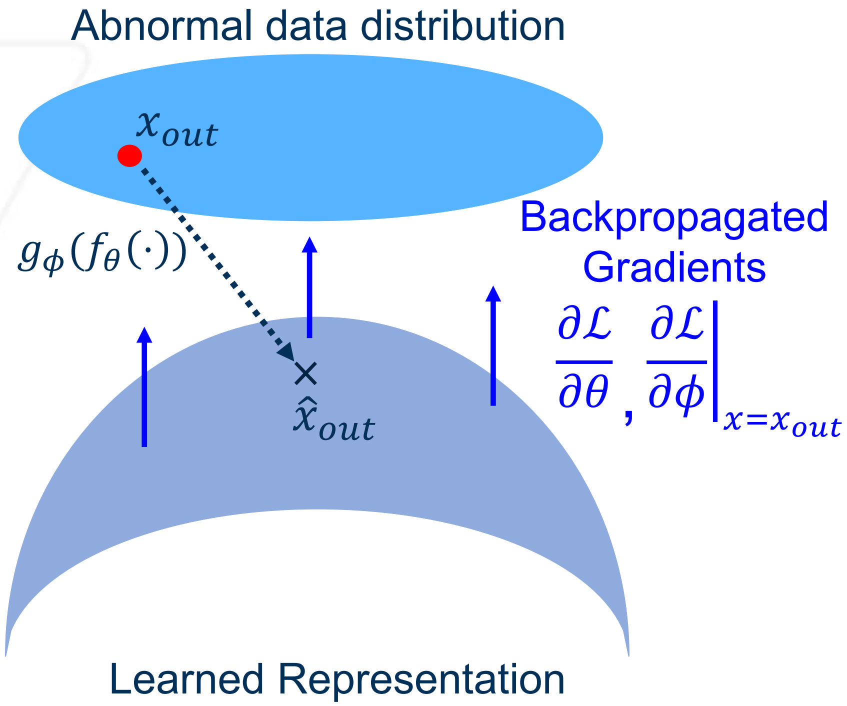


Gawlikowski, J., Tassi, C. R. N., Ali, M., Lee, J., Humt, M., Feng, J., ... & Zhu, X. X. (2021). A survey of uncertainty in deep neural networks. *arXiv preprint arXiv:2107.03342*.

Probing the Purview of Neural Networks via Gradient Analysis

SCAN ME

Principle: Gradients provide a distance measure between the learned representations space and novel data



However, what is $\mathcal{L}$?

- In anomaly detection, the loss was between the input and its reconstruction
- In prediction tasks, there is neither the reconstructed input or ground truth

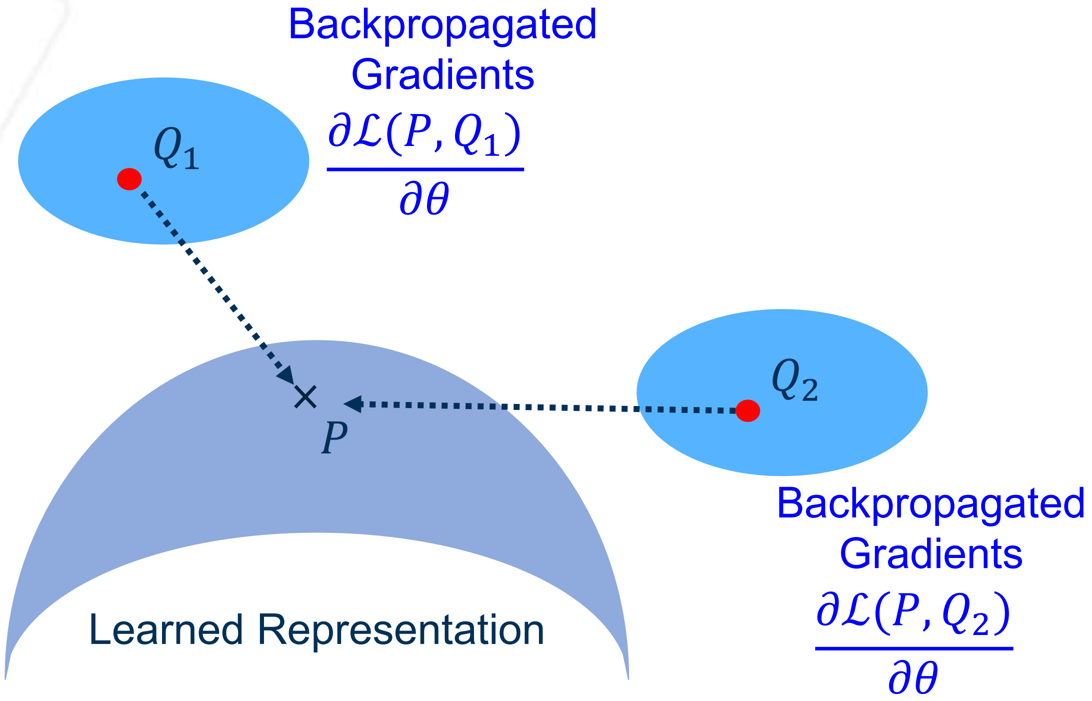
Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

SCAN ME

Probing the Purview of Neural Networks via Gradient Analysis

Principle: Gradients provide a distance measure between the learned representations space and novel data

P = Predicted class
Q_1 = Contrast class 1
Q_2 = Contrast class 2



However, what is $\mathcal{L}$?

- In anomaly detection, the loss was between the input and its reconstruction
- In prediction tasks, there is neither the reconstructed input or ground truth

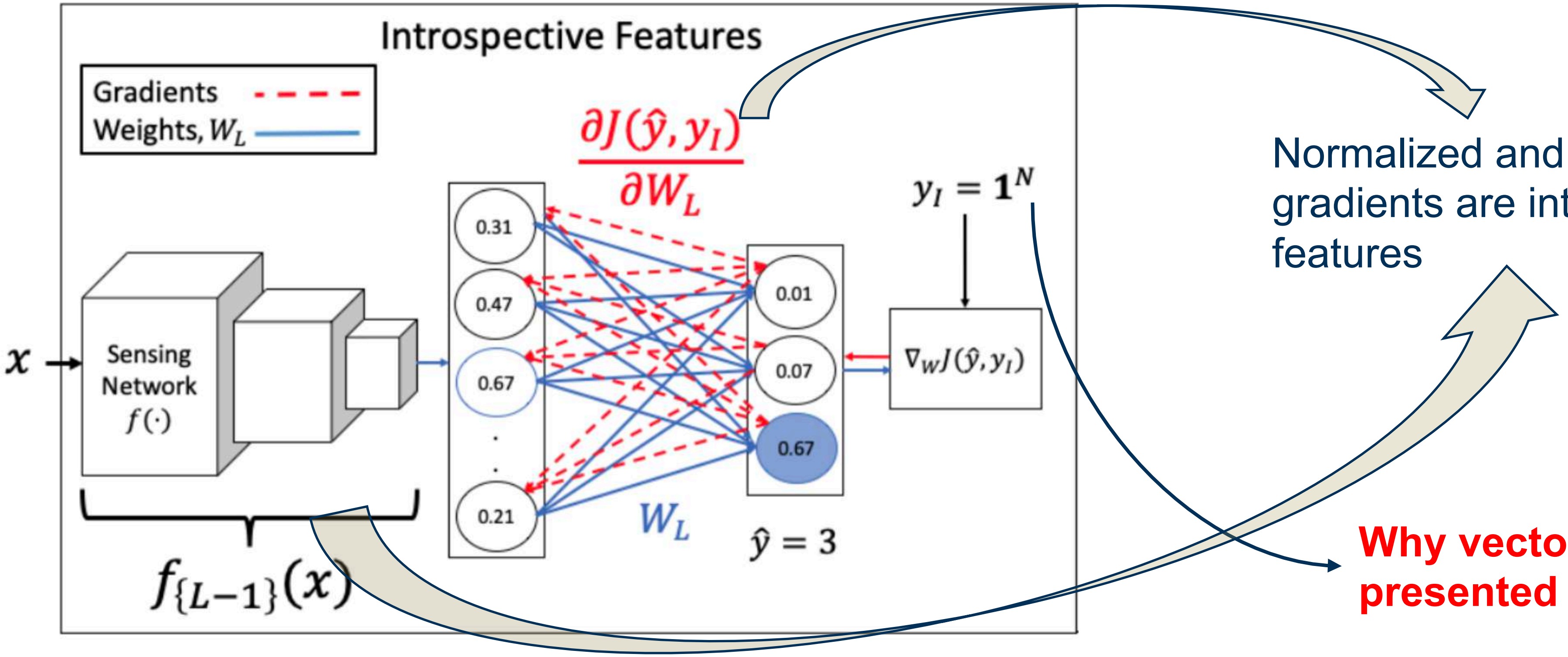
- **We backpropagate all possible classes - $Q_1, Q_2 \ldots Q_N$ by backpropagating N one-hot vectors**
- Higher the distance to all classes, higher the uncertainty score

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

SCAN ME

Probing the Purview of Neural Networks via Gradient Analysis

Step 1: Measure the loss between the prediction P and a vector of all ones and backpropagate to obtain the introspective features



Normalized and vectorized gradients are introspective features

Why vector of all 1s? The theory is presented in [1]

[1] M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

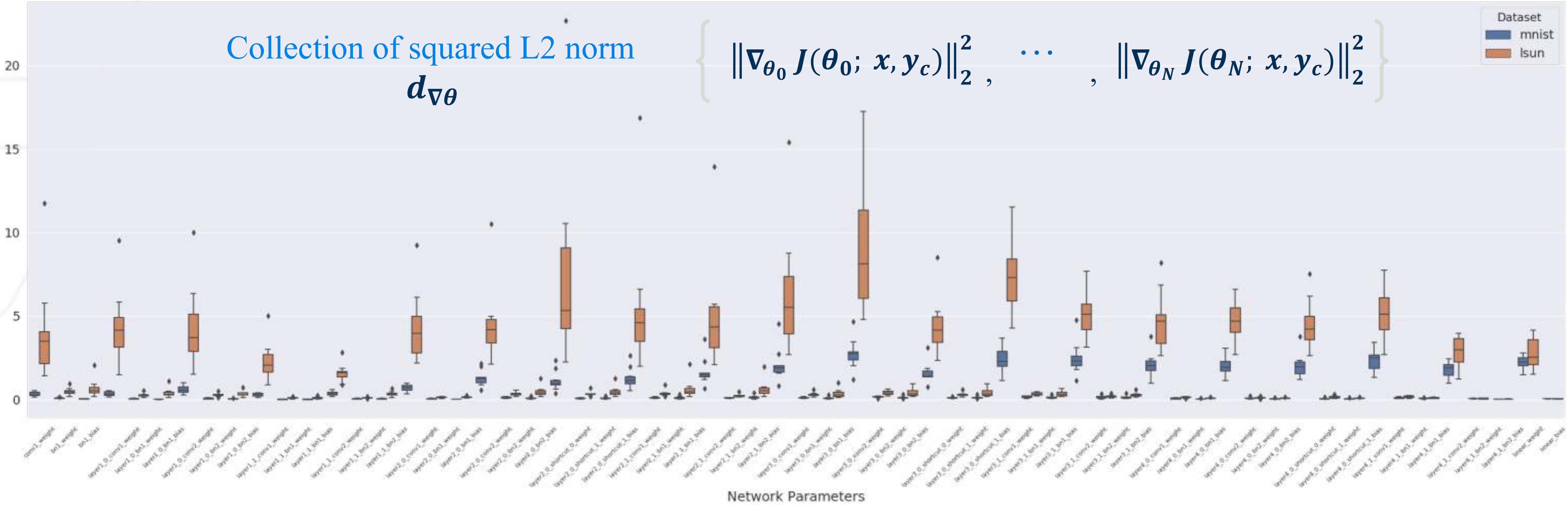
SCAN ME

Probing the Purview of Neural Networks via Gradient Analysis

Step 2: Take L2 norm of all generated gradients

Collection of squared L2 norm $d_{\nabla\theta}$

$$\left\{ \left\|\nabla_{\theta_0} J(\theta_0;\, x, y_c)\right\|_2^2,\ \cdots\ ,\ \left\|\nabla_{\theta_N} J(\theta_N;\, x, y_c)\right\|_2^2 \right\}$$



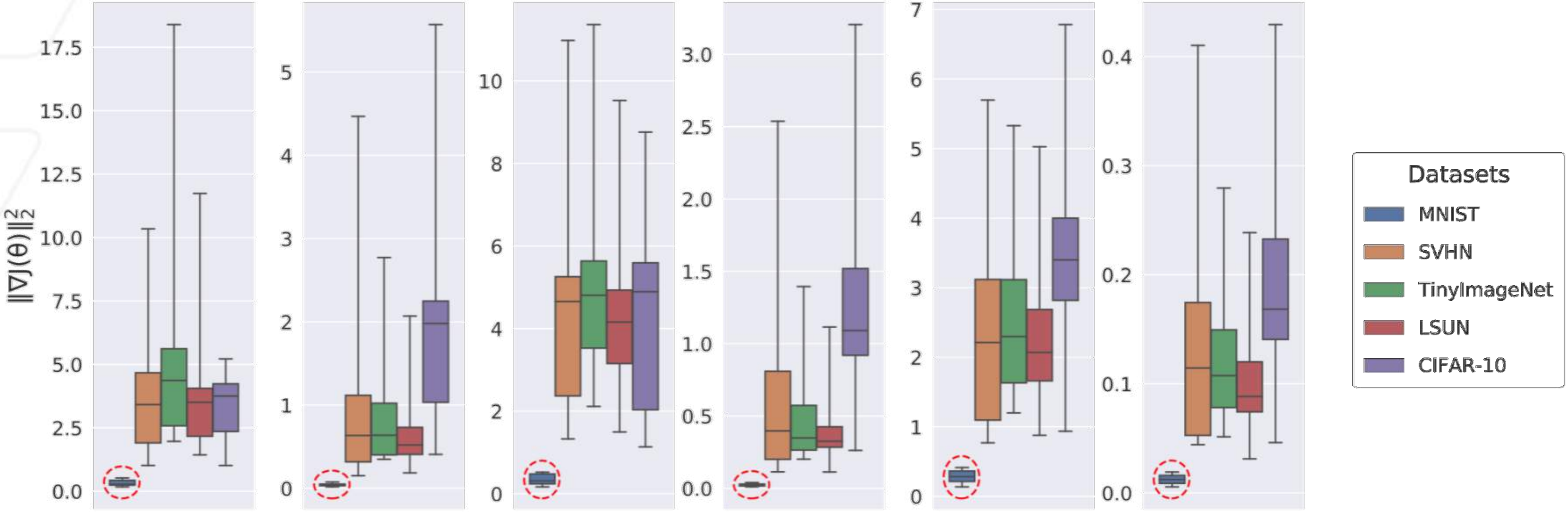
MNIST: In-distribution, SUN: Out-of-Distribution

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Gradient-based Uncertainty
Uncertainty Results in OOD setting

Probing the Purview of Neural Networks via Gradient Analysis

Squared L2 distances for different parameter sets



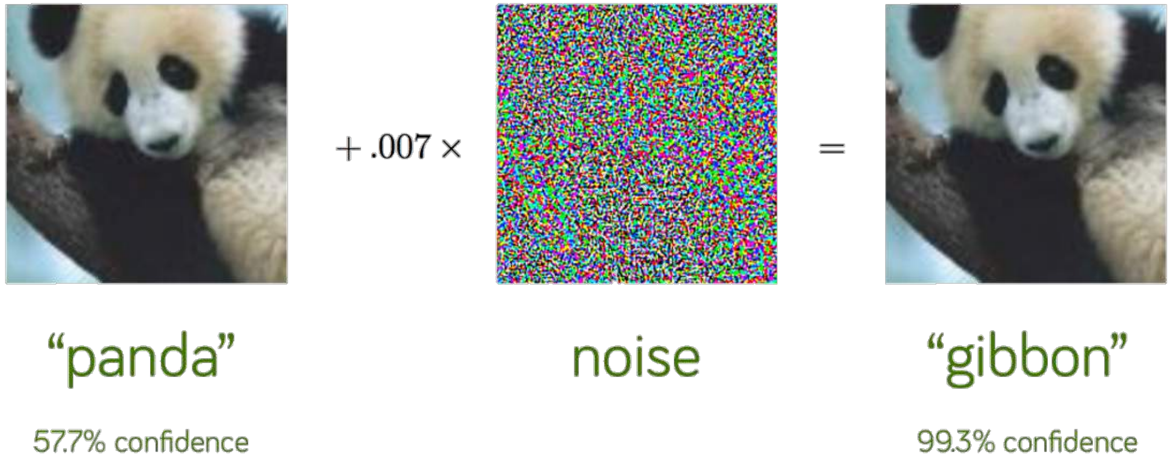
MNIST: Circled in red. Significantly lower uncertainty compared to OOD datasets

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Gradient-based Uncertainty
Uncertainty Results in Adversarial Setting

Probing the Purview of Neural Networks via Gradient Analysis

Vulnerable DNNs in the real world



Goal: to examine the ability of trained DNNs to handle adversarial inputs during inference

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Gradient-based Uncertainty
Uncertainty Results in Adversarial Setting

SCAN ME

Probing the Purview of Neural Networks via Gradient Analysis

Model	Attacks	Baseline	LID	M(V)	M(P)	M(FE)	M(P+FE)	Ours
ResNet	FGSM	51.20	90.06	81.69	84.25	**99.95**	**99.95**	93.45
	BIM	49.94	99.21	87.09	89.20	**100.0**	**100.0**	96.19
	C&W	53.40	76.47	74.51	75.71	92.78	92.79	**97.07**
	PGD	50.03	67.48	56.27	57.57	65.23	75.98	**95.82**
	IterLL	60.40	85.17	62.32	64.10	85.10	92.10	**98.17**
	Semantic	52.29	86.25	64.18	65.79	83.95	84.38	**90.15**
DenseNet	FGSM	52.76	98.23	86.88	87.24	**99.98**	99.97	96.83
	BIM	49.67	**100.0**	89.19	89.17	**100.0**	**100.0**	96.85
	C&W	54.53	80.58	75.77	76.16	90.83	90.76	**97.05**
	PGD	49.87	83.01	70.39	66.52	86.94	83.61	**96.77**
	IterLL	55.43	83.16	70.17	66.61	83.20	77.84	**98.53**
	Semantic	53.54	81.41	62.16	62.15	67.98	67.29	**89.55**

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

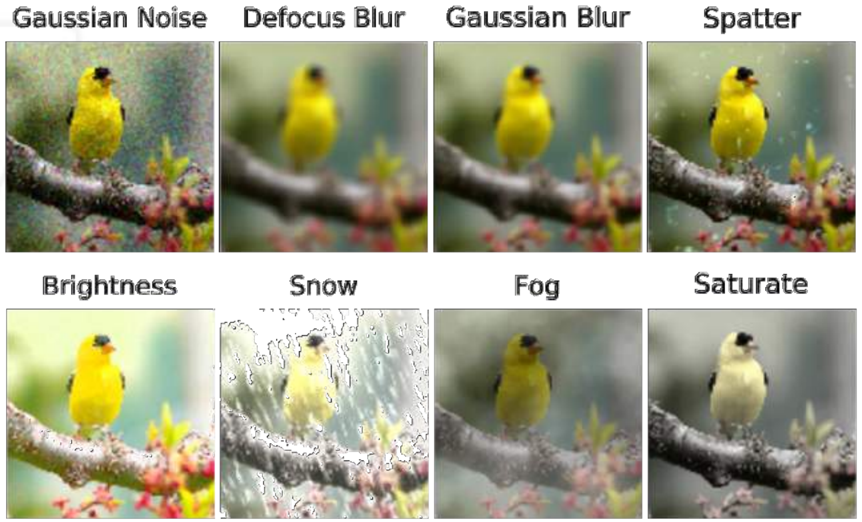
Gradient-based Uncertainty

Uncertainty Results to Detect Challenging Conditions

SCAN ME

Same application as Anomaly Detection, except there is no need for an additional AE network!

CIFAR-10-C



CURE-TSR

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Gradient-based Uncertainty

Uncertainty Results to Detect Challenging Conditions

Probing the Purview of Neural Networks via Gradient Analysis

SCAN ME

Dataset	Method	Mahalanobis [12] / Ours				
	Corruption	Level 1	Level 2	Level 3	Level 4	Level 5
CIFAR-10-C	Noise	96.63 / **99.95**	98.73 / **99.97**	99.46 / **99.99**	99.62 / **99.97**	99.71 / **99.99**
	LensBlur	94.22 / **99.95**	97.51 / **99.99**	99.26 / **100.0**	99.78 / **100.0**	99.89 / **100.0**
	GaussianBlur	94.19 / **99.94**	99.28 / **100.0**	99.76 / **100.0**	99.86 / **100.0**	99.80 / **100.0**
	DirtyLens	93.37 / **99.94**	95.31 / **99.93**	95.66 / **99.96**	95.37 / **99.92**	97.43 / **99.96**
	Exposure	91.39 / **99.87**	91.00 / **99.85**	90.71 / **99.88**	90.58 / **99.85**	90.68 / **99.87**
	Snow	93.64 / **99.94**	96.50 / **99.94**	94.44 / **99.95**	94.22 / **99.95**	95.25 / **99.92**
	Haze	95.52 / **99.95**	98.35 / **99.99**	99.28 / **100.0**	99.71 / **99.99**	99.94 / **100.0**
	Decolor	93.51 / **99.96**	93.55 / **99.96**	90.30 / **99.82**	89.86 / **99.75**	90.43 / **99.83**
CURE-TSR	Noise	25.46 / **50.20**	47.54 / **63.87**	47.32 / **81.20**	66.19 / **91.16**	83.14 / **94.81**
	LensBlur	48.06 / **72.63**	71.61 / **87.58**	86.59 / **92.56**	92.19 / **93.90**	94.90 / **95.65**
	GaussianBlur	66.44 / **83.07**	77.67 / **86.94**	93.15 / **94.35**	80.78 / **94.51**	**97.36** / 96.53
	DirtyLens	29.78 / **51.21**	29.28 / **59.10**	46.60 / **82.10**	73.36 / **91.87**	98.50 / **98.70**
	Exposure	74.90 / **88.13**	**99.96** / 96.78	**99.99** / 99.26	**100.0** / 99.80	**100.0** / 99.90
	Snow	28.11 / **61.34**	61.28 / **80.52**	89.89 / **91.30**	**99.34** / 96.13	**99.98** / 97.66
	Haze	66.51 / **95.83**	97.86 / **99.50**	**100.0** / 99.95	**100.0** / 99.87	**100.0** / 99.88
	Decolor	48.37 / **62.36**	60.55 / **81.30**	71.73 / **89.93**	87.29 / **95.42**	89.68 / **96.91**

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Gradient-based Uncertainty

Uncertainty Results to Detect Challenging Conditions

Probing the Purview of Neural Networks via Gradient Analysis

SCAN ME

Dataset	Method / Corruption	Mahalanobis [12] / Ours Level 1	Level 2	Level 3	Level 4	Level 5
CIFAR-10-C	Noise	96.63 / **99.95**	98.73 / **99.97**	99.46 / **99.99**	99.62 / **99.97**	99.71 / **99.99**
	LensBlur	94.22 / **99.95**	97.51 / **99.99**	99.26 / **100.0**	99.78 / **100.0**	99.89 / **100.0**
	GaussianBlur	94.19 / **99.94**	99.28 / **100.0**	99.76 / **100.0**	99.86 / **100.0**	99.80 / **100.0**
	DirtyLens	93.37 / **99.94**	95.31 / **99.93**	95.66 / **99.96**	95.37 / **99.92**	97.43 / **99.96**
	Exposure	91.39 / **99.87**	91.00 / **99.85**	90.71 / **99.88**	90.58 / **99.85**	90.68 / **99.87**
	Snow	93.64 / **99.94**	96.50 / **99.94**	94.44 / **99.95**	94.22 / **99.95**	95.25 / **99.92**
	Haze	95.52 / **99.95**	98.35 / **99.99**	99.28 / **100.0**	99.71 / **99.99**	99.94 / **100.0**
	Decolor	93.51 / **99.96**	93.55 / **99.96**	90.30 / **99.82**	89.86 / **99.75**	90.43 / **99.83**
CURE-TSR	Noise	25.46 / **50.20**	47.54 / **63.87**	47.32 / **81.20**	66.19 / **91.16**	83.14 / **94.81**
	LensBlur	48.06 / **72.63**	71.61 / **87.58**	86.59 / **92.56**	92.19 / **93.90**	94.90 / **95.65**
	GaussianBlur	66.44 / **83.07**	77.67 / **86.94**	93.15 / **94.35**	80.78 / **94.51**	**97.36** / 96.53
	DirtyLens	29.78 / **51.21**	29.28 / **59.10**	46.60 / **82.10**	73.36 / **91.87**	98.50 / **98.70**
	Exposure	74.90 / **88.13**	**99.96** / 96.78	**99.99** / 99.26	**100.0** / 99.80	**100.0** / 99.90
	Snow	28.11 / **61.34**	61.28 / **80.52**	89.89 / **91.30**	**99.34** / 96.13	**99.98** / 97.66
	Haze	66.51 / **95.83**	97.86 / **99.50**	**100.0** / 99.95	**100.0** / 99.87	**100.0** / 99.88
	Decolor	48.37 / **62.36**	60.55 / **81.30**	71.73 / **89.93**	87.29 / **95.42**	89.68 / **96.91**

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

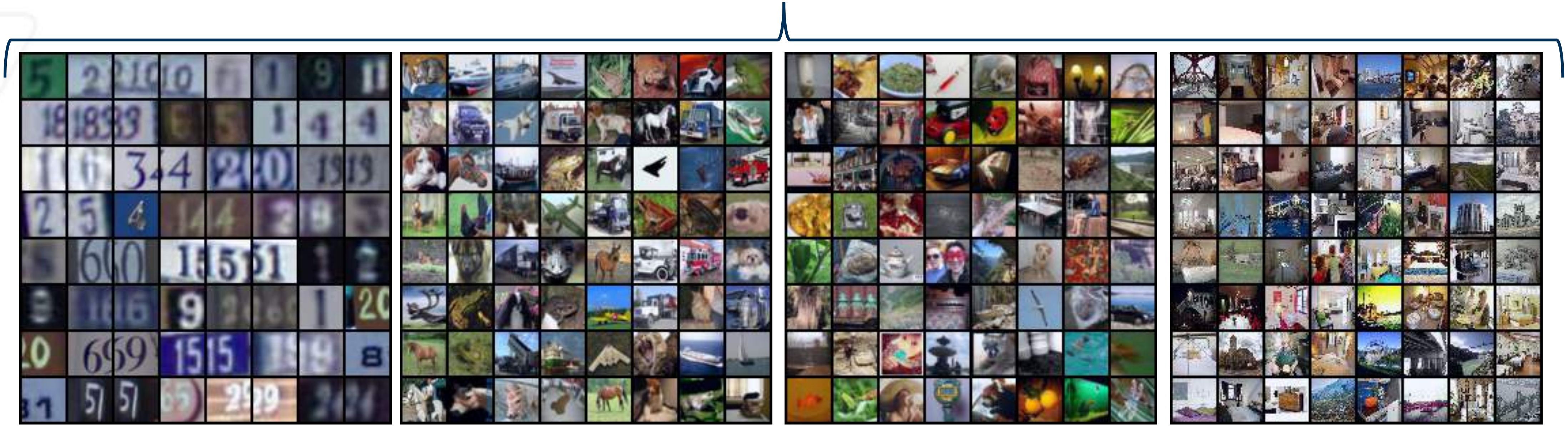
Out-of-Distribution Detection

Probing the Purview of Neural Networks via Gradient Analysis

SCAN ME

Train set →  MNIST

Goal: To detect that these datasets are not part of training



SVHN | CIFAR10 | TinyImageNet | LSUN

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Out-of-Distribution Detection

Probing the Purview of Neural Networks via Gradient Analysis

SCAN ME

Dataset Distribution In	Out	Detection Accuracy	AUROC	AUPR
		Baseline [5] / ODIN [6] / Mahalanobis (V) [7] / Mahalanobis (P+FE) [7] / Ours		
CIFAR-10	SVHN	83.36 / 88.81 / 79.39 / 91.95 / **98.04**	88.30 / 94.93 / 85.03 / 97.10 / **99.84**	88.26 / 95.45 / 86.15 / 96.12 / **99.98**
	TinyImageNet	84.01 / 85.21 / 83.60 / **97.45** / 86.17	90.06 / 91.86 / 88.93 / **99.68** / 93.18	89.26 / 91.60 / 88.59 / **99.60** / 92.66
	LSUN	87.34 / 88.42 / 85.02 / **98.60** / 98.37	92.79 / 94.48 / 90.11 / **99.86** / **99.86**	92.30 / 94.22 / 89.80 / 99.82 / **99.87**
SVHN	CIFAR-10	79.98 / 80.12 / 74.10 / 88.84 / **97.90**	81.50 / 81.49 / 79.31 / 95.05 / **99.79**	81.01 / 80.95 / 80.83 / 90.25 / **98.11**
	TinyImageNet	81.70 / 81.92 / 79.35 / 96.17 / **97.74**	83.69 / 83.82 / 83.85 / 99.23 / **99.77**	82.54 / 82.60 / 85.50 / **98.17** / 97.93
	LSUN	80.96 / 81.15 / 79.52 / 97.50 / **99.04**	82.85 / 82.98 / 83.02 / 99.54 / **99.93**	81.97 / 82.01 / 84.67 / 98.84 / **99.21**

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Out-of-Distribution Detection

SCAN ME

Probing the Purview of Neural Networks via Gradient Analysis

Dataset Distribution		Detection Accuracy	AUROC	AUPR
In	Out	Baseline [5] / ODIN [6] / Mahalanobis (V) [7] / Mahalanobis (P+FE) [7] / Ours		
CIFAR-10	SVHN	83.36 / 88.81 / 79.39 / 91.95 / **98.04**	88.30 / 94.93 / 85.03 / 97.10 / **99.84**	88.26 / 95.45 / 86.15 / 96.12 / **99.98**
	TinyImageNet	84.01 / 85.21 / 83.60 / **97.45** / 86.17	90.06 / 91.86 / 88.93 / **99.68** / 93.18	89.26 / 91.60 / 88.59 / **99.60** / 92.66
	LSUN	87.34 / 88.42 / 85.02 / **98.60** / 98.37	92.79 / 94.48 / 90.11 / **99.86** / **99.86**	92.30 / 94.22 / 89.80 / 99.82 / **99.87**
SVHN	CIFAR-10	79.98 / 80.12 / 74.10 / 88.84 / **97.90**	81.50 / 81.49 / 79.31 / 95.05 / **99.79**	81.01 / 80.95 / 80.83 / 90.25 / **98.11**
	TinyImageNet	81.70 / 81.92 / 79.35 / 96.17 / **97.74**	83.69 / 83.82 / 83.85 / 99.23 / **99.77**	82.54 / 82.60 / 85.50 / **98.17** / 97.93
	LSUN	80.96 / 81.15 / 79.52 / 97.50 / **99.04**	82.85 / 82.98 / 83.02 / 99.54 / **99.93**	81.97 / 82.01 / 84.67 / 98.84 / **99.21**

Numbers

SVHN

CIFAR10

TinyImageNet

LSUN

Objects, natural scenes

IEEE IV 2023 IEEE INTELLIGENT VEHICLES SYMPOSIUM

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

OLIVES @GeorgiaTech

Georgia Tech

Out-of-Distribution Detection

Probing the Purview of Neural Networks via Gradient Analysis

SCAN ME

Dataset Distribution		Detection Accuracy	AUROC	AUPR
In	Out	Baseline [5] / ODIN [6] / Mahalanobis (V) [7] / Mahalanobis (P+FE) [7] / Ours		
CIFAR-10	SVHN	83.36 / 88.81 / 79.39 / 91.95 / **98.04**	88.30 / 94.93 / 85.03 / 97.10 / **99.84**	88.26 / 95.45 / 86.15 / 96.12 / **99.98**
	TinyImageNet	84.01 / 85.21 / 83.60 / **97.45** / 86.17	90.06 / 91.86 / 88.93 / **99.68** / 93.18	89.26 / 91.60 / 88.59 / **99.60** / 92.66
	LSUN	87.34 / 88.42 / 85.02 / **98.60** / 98.37	92.79 / 94.48 / 90.11 / **99.86** / **99.86**	92.30 / 94.22 / 89.80 / 99.82 / **99.87**
SVHN	CIFAR-10	79.98 / 80.12 / 74.10 / 88.84 / **97.90**	81.50 / 81.49 / 79.31 / 95.05 / **99.79**	81.01 / 80.95 / 80.83 / 90.25 / **98.11**
	TinyImageNet	81.70 / 81.92 / 79.35 / 96.17 / **97.74**	83.69 / 83.82 / 83.85 / 99.23 / **99.77**	82.54 / 82.60 / 85.50 / **98.17** / 97.93
	LSUN	80.96 / 81.15 / 79.52 / 97.50 / **99.04**	82.85 / 82.98 / 83.02 / 99.54 / **99.93**	81.97 / 82.01 / 84.67 / 98.84 / **99.21**

More similar datasets (objects)

TinyImageNet

CIFAR10

LSUN

SVHN

Lee, Jinsol, et al. "Probing the Purview of Neural Networks via Gradient Analysis." *IEEE Access* 11 (2023): 32716-32732.

Inference

Overcoming Deficiencies at Inference

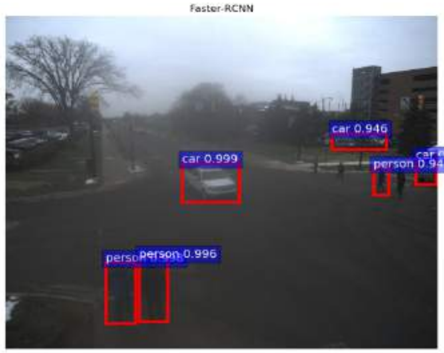
What is required when networks are met with challenging data at inference?

To overcome deficiencies, predictions from neural networks must be equipped with:

- Anomaly scores: How *close* to the training data is the novel data at inference?
- Uncertainty scores: How close to the *best* possible network is the trained network?
- **Contextual Explainability**: How *relevant* are the network explanations for its prediction?

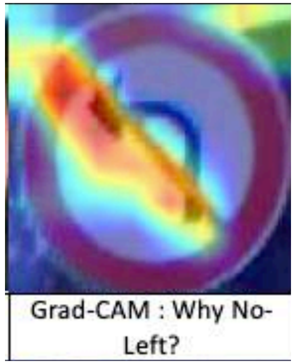
Training data

Anomalous data



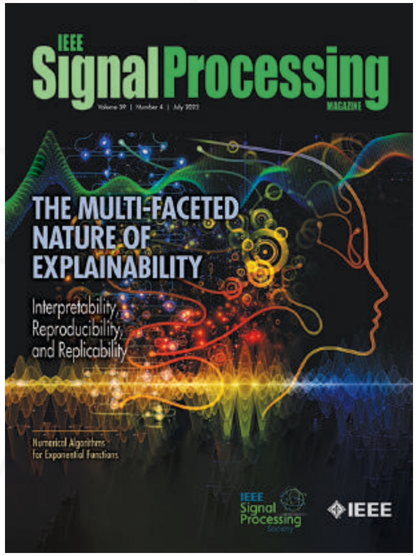
Certain objects

Uncertain objects



'Why P'

'Why P, rather than Q?'



Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

Mohit Prabhushankar, PhD
Postdoc

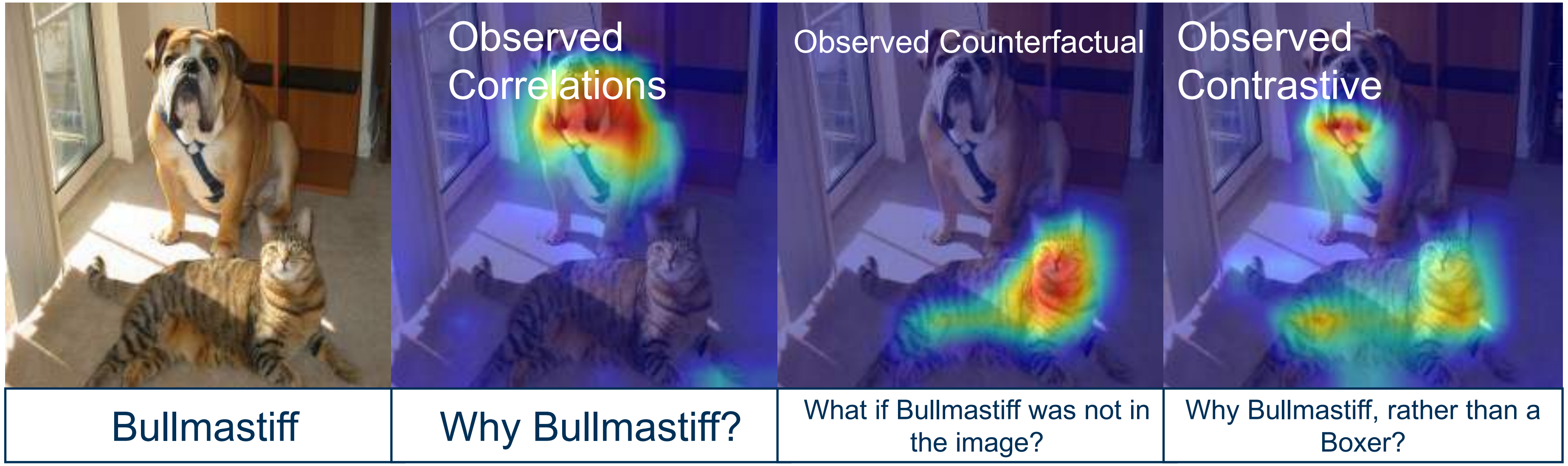
Ghassan AlRegib, PhD
Professor

SCAN ME

What are Visual Explanations?

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

- **Explanations are defined as a set of rationales used to understand the reasons behind a decision**
- **If the decision is based on visual characteristics within the data, the decision-making reasons are visual explanations**



AlRegib, G., & Prabhushankar, M. (2022). Explanatory Paradigms in Neural Networks: Towards relevant and contextual explanations. *IEEE Signal Processing Magazine*, *39*(4), 59-72.

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

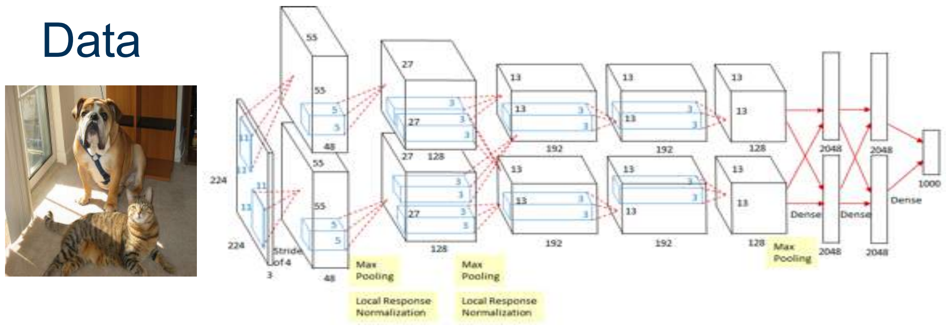
Explainability matters establishes trust in deep learning systems by developing *transparent* models that can explain *why they predict what they predict* to humans

Explainability is useful in*:*

- Medical: help doctors diagnose
- Seismic: help interpreters label seismic data
- Autonomous Systems: build appropriate trust and confidence

Algorithm

Data



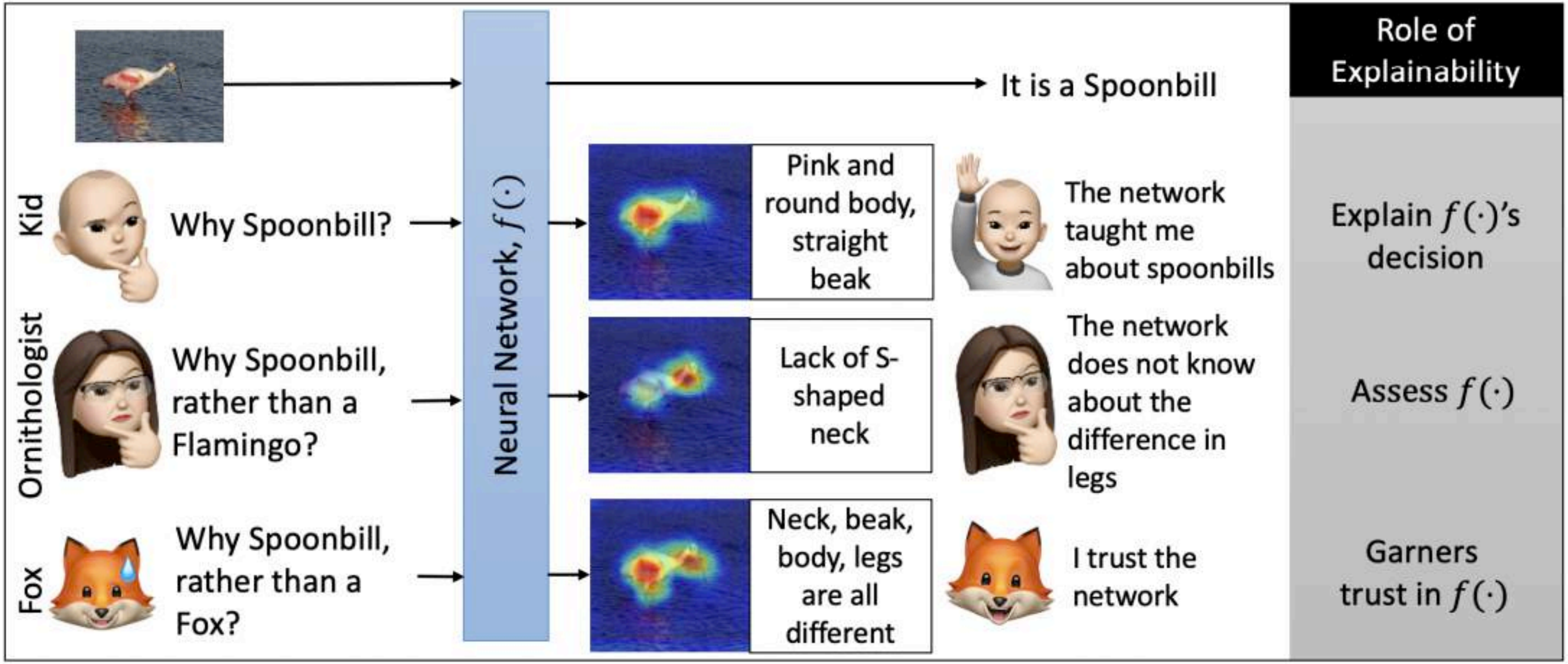
Output

class scores

Deep models act as algorithms that take data and output something **without** being able to **explain** their methodology

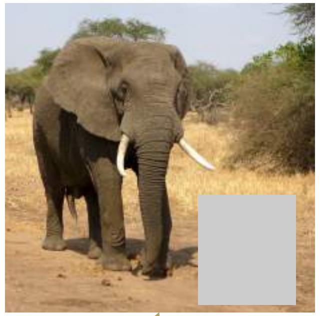
AlRegib, G., & Prabhushankar, M. (2022). Explanatory Paradigms in Neural Networks: Towards relevant and contextual explanations. *IEEE Signal Processing Magazine*, *39*(4), 59-72.

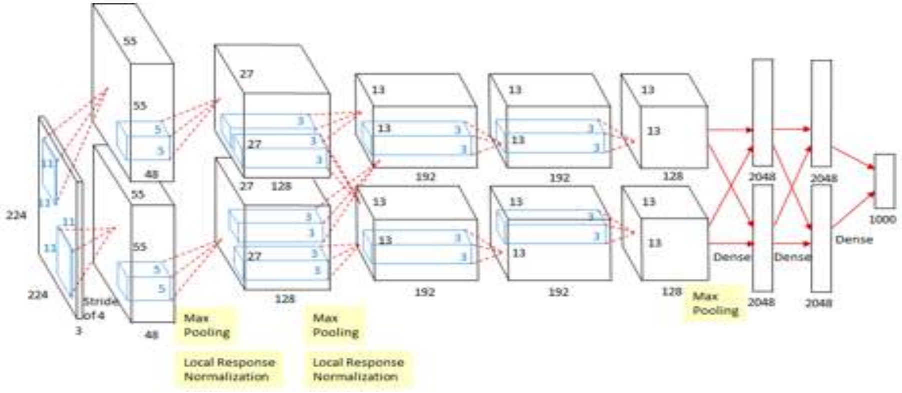
Explanations
Role of Visual Explanations



AlRegib, G., & Prabhushankar, M. (2022). Explanatory Paradigms in Neural Networks: Towards relevant and contextual explanations. *IEEE Signal Processing Magazine*, *39*(4), 59-72.

Intervention: Mask part of the image before feeding to CNN, check how much predicted probabilities change





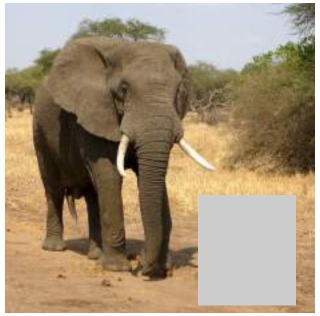
P(elephant) = 0.95

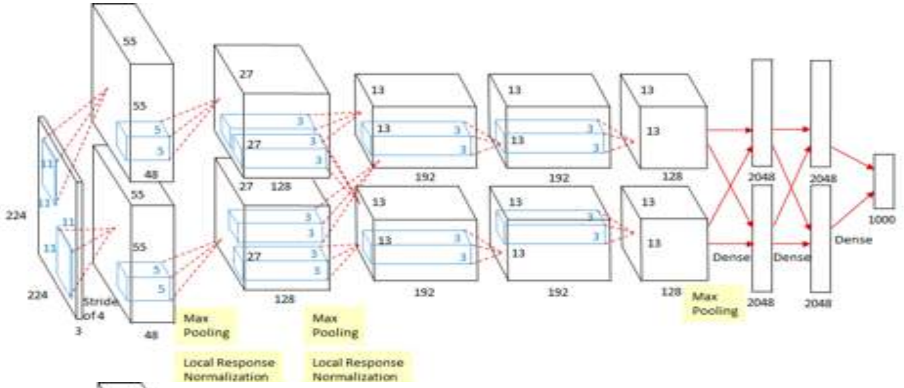
A gray patch or patch of average pixel value of the dataset
Note: not a black patch because the input images are centered to zero in the preprocessing.

Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, ECCV 2014

Explanations
Input Saliency via Occlusion

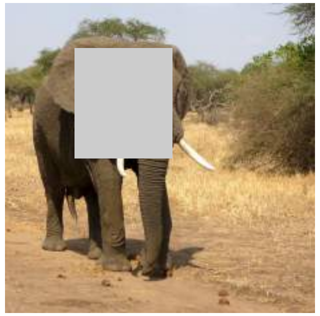
Intervention: Mask part of the image before feeding to CNN, check how much predicted probabilities change

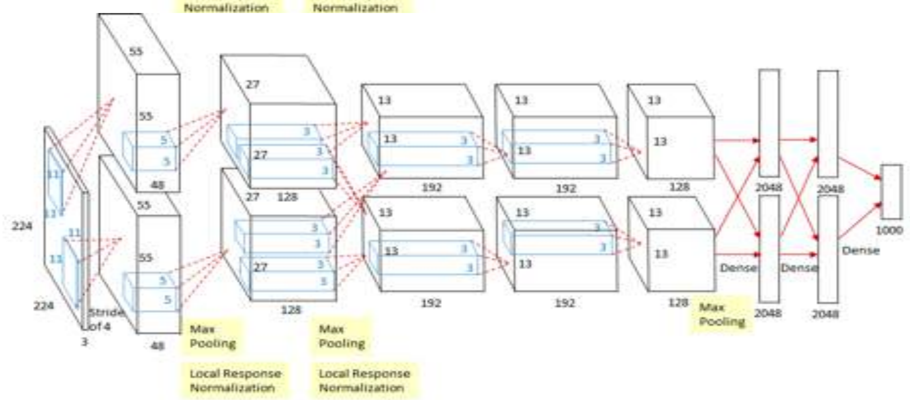




P(elephant) = 0.95

These pixels affect decisions more





P(elephant) = 0.75

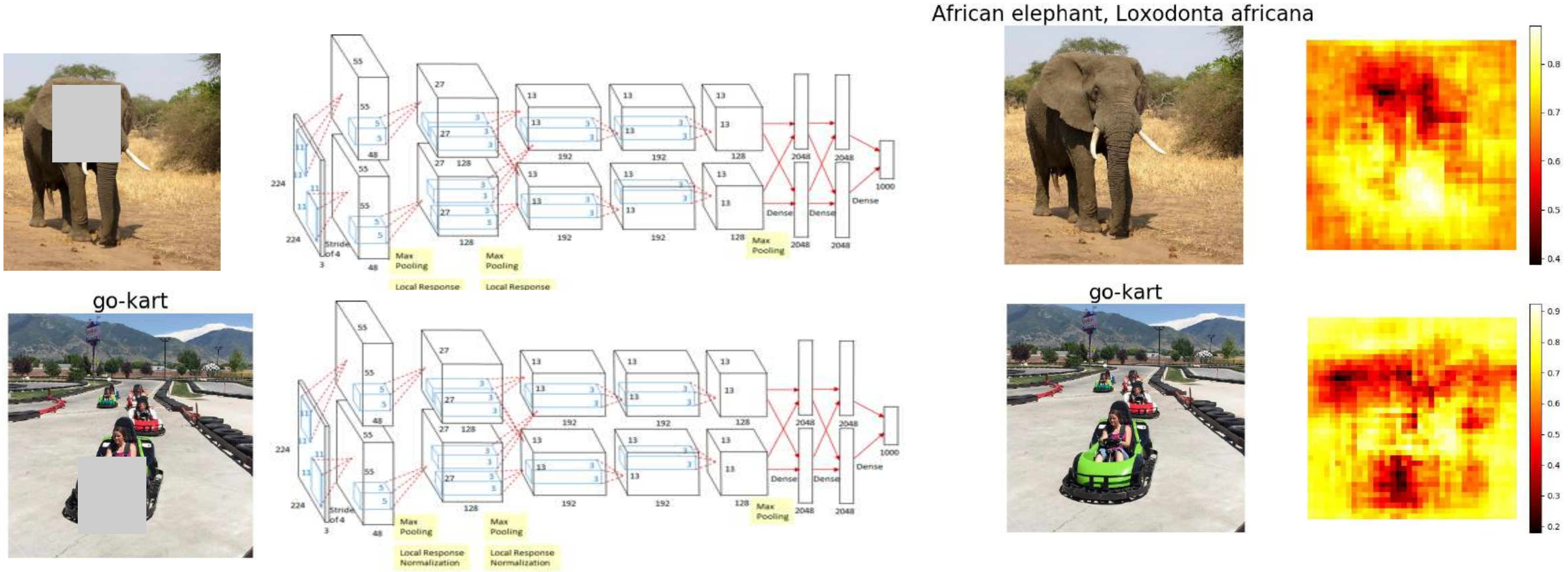
Zeiler and Fergus, "Visualizing and Understanding Convolutional Networks", ECCV 2014

Explanations
Input Saliency via Occlusion

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

The network is trained with image- labels, but it is sensitive to the common visual regions in images



Zeiler and Fergus, “Visualizing and Understanding Convolutional Networks”, ECCV 2014

Explanations
Input Saliency via Gradients

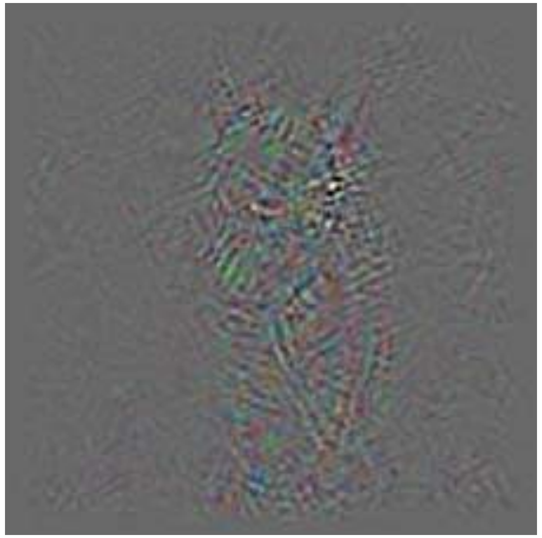
Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

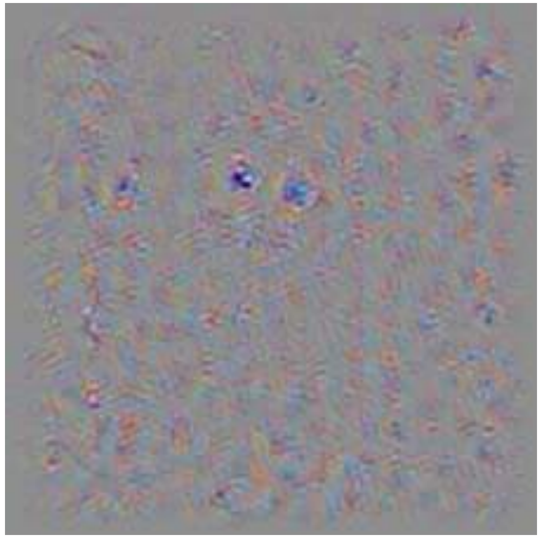
Gradients provide a one-shot means of perturbing the input that changes the output

Input

Vanilla Gradients



Deconvolution Gradients



Guided Backpropagation

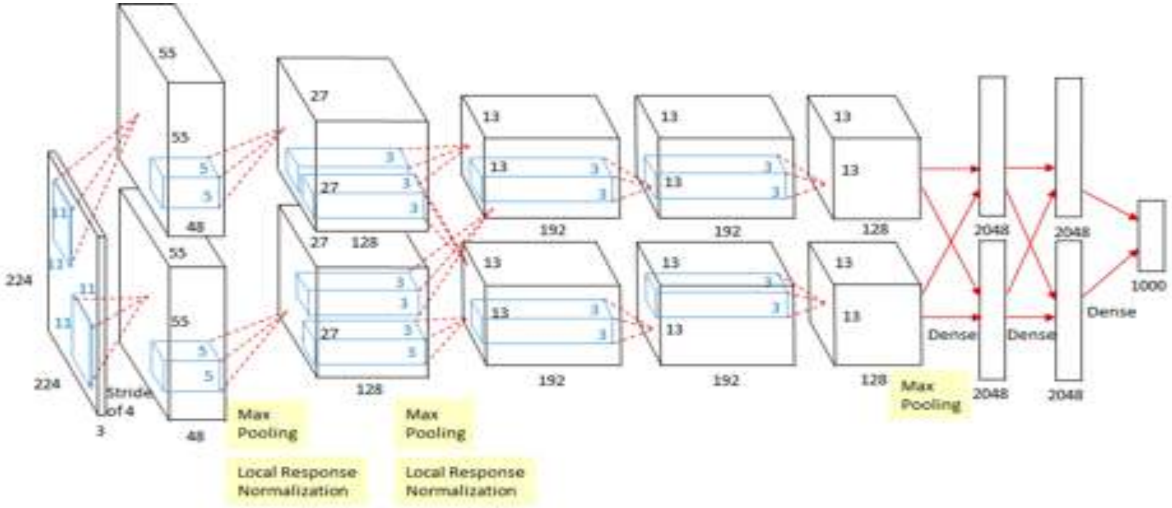
However, localization remains an issue

Springenberg, Dosovitskiy, et al., Striving for Simplicity: The all convolutional net, 2015

Gradient and Activation-based Explanations
GradCAM

Gradients provide a one-shot means of perturbing the input that changes the output. Activations provide the localization.

- To find the important activations that are responsible for a particular class
- We want the activations:
  - **Class-discriminative** to reflect decision-making
  - **Preserve spatial information** to ensure spatial coverage of important regions



Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Gradient and Activation-based Explanations
GradCAM

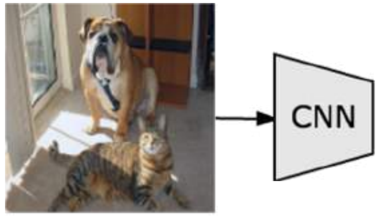
Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

Gradients provide a one-shot means of perturbing the input that changes the output. Activations provide the localization.

- Given an image, feed forward through CNN

image

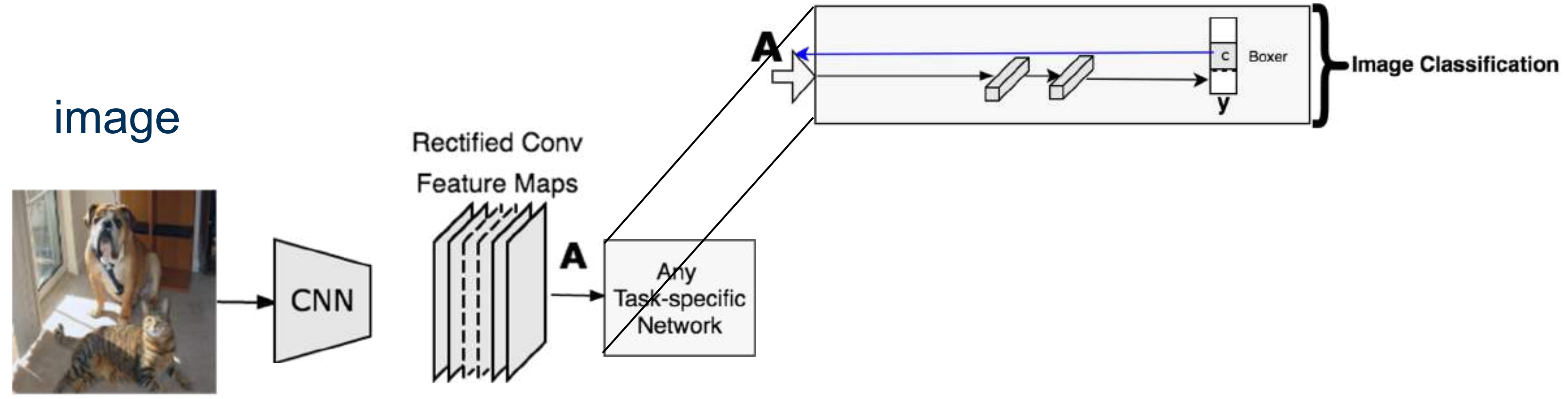


Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Gradient and Activation-based Explanations
GradCAM

Gradients provide a one-shot means of perturbing the input that changes the output. Activations provide the localization.

- Given an image, feed forward through CNN
- Final convolutional layer output feature maps for later task-specific layers, i.e., fc layer for classification



Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

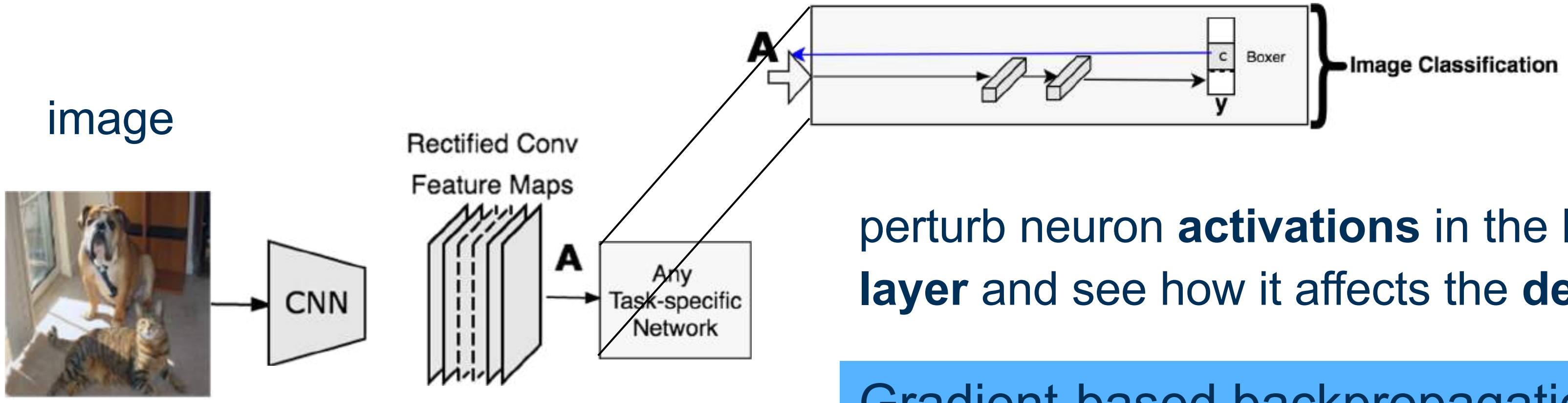
Gradient and Activation-based Explanations
GradCAM

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

Gradients provide a one-shot means of perturbing the input that changes the output. Activations provide the localization.

- Given an image, feed forward through CNN
- Final convolutional layer output feature maps for later task-specific layers, i.e., fc layer for classification



perturb neuron **activations** in the **last conv layer** and see how it affects the **decision**

Gradient-based backpropagation to obtain activation importance

Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

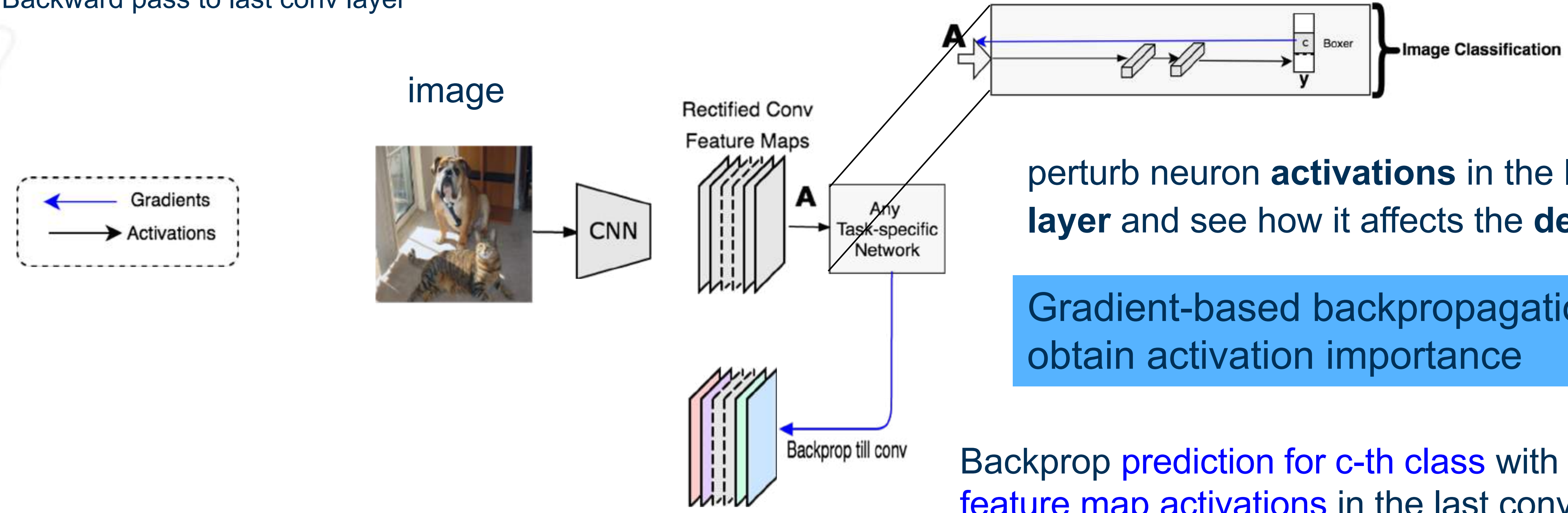
Gradient and Activation-based Explanations
GradCAM

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

Gradients provide a one-shot means of perturbing the input that changes the output. Activations provide the localization.

- Given an image, feed forward through CNN
- Final convolutional layer output feature maps for later task-specific layers, i.e., fc layer for classification
- Backward pass to last conv layer



perturb neuron **activations** in the **last conv layer** and see how it affects the **decision**

Gradient-based backpropagation to obtain activation importance

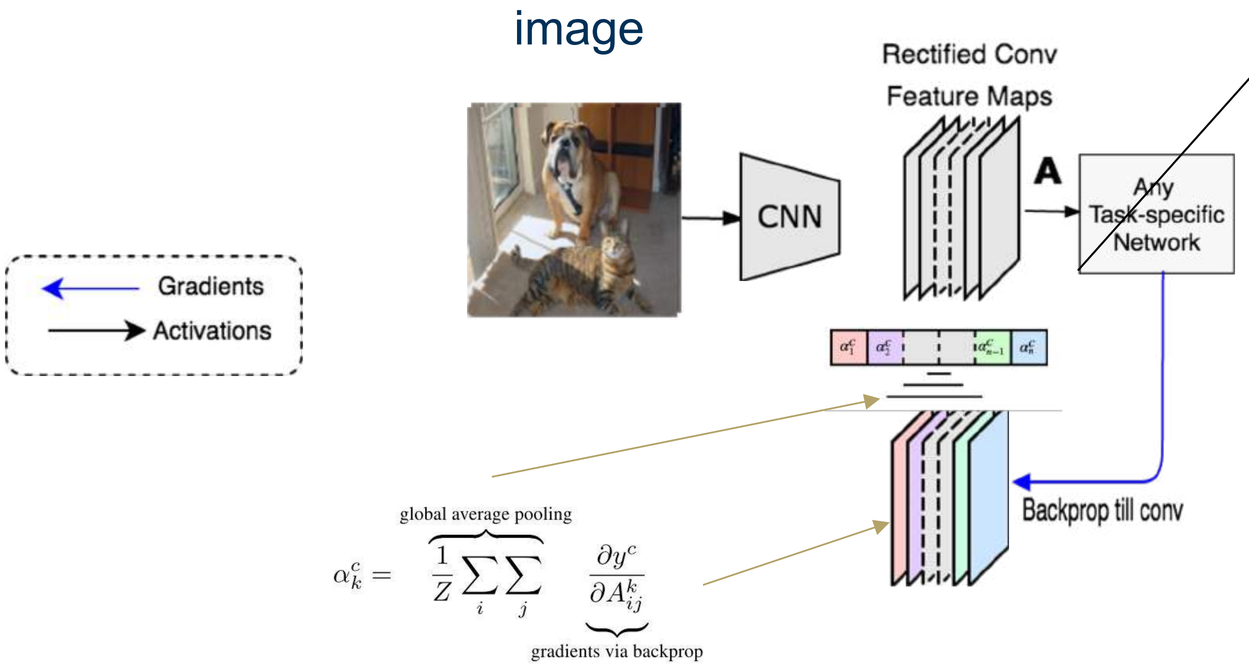
Backprop prediction for c-th class with respect to feature map activations in the last conv layer

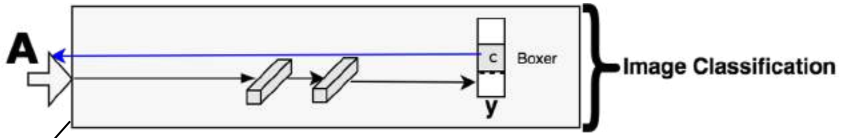
Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Gradient and Activation-based Explanations
GradCAM

Gradients provide a one-shot means of perturbing the input that changes the output. Activations provide the localization.

- Given an image, feed forward through CNN
- Final convolutional layer output feature maps for later task-specific layers, i.e., fc layer for classification
- Backward pass to last conv layer
- Compute gradients w.r.t. last conv activations





$\frac{\partial y^c}{\partial A^k}$: gradients of prediction for c-th class with respect to k-th feature map activations A^k in the last conv layer

α_k^c is the scalar importance of k-th feature map obtained by averaging $\frac{\partial y^c}{\partial A^k}$ spatially

Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

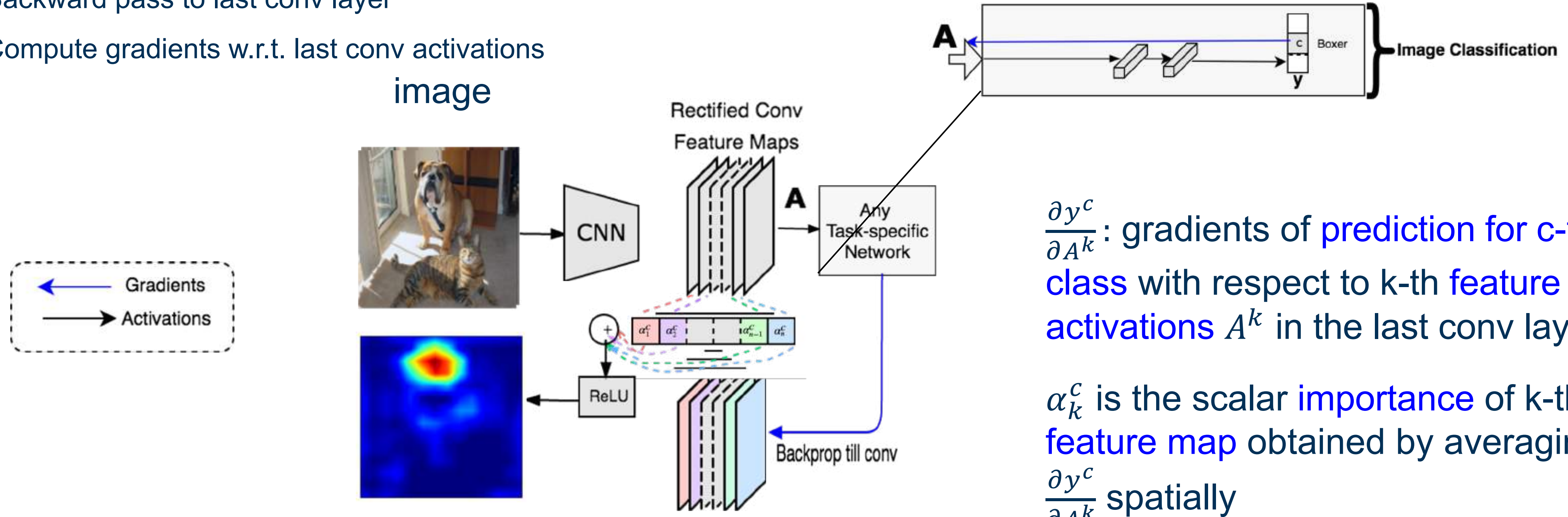
Gradient and Activation-based Explanations
GradCAM

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

Gradients provide a one-shot means of perturbing the input that changes the output. Activations provide the localization.

- Given an image, feed forward through CNN
- Final convolutional layer output feature maps for later task-specific layers, i.e., fc layer for classification
- Backward pass to last conv layer
- Compute gradients w.r.t. last conv activations



Grad-CAM (up-sampled to original image dimension)

$\frac{\partial y^c}{\partial A^k}$: gradients of prediction for c-th class with respect to k-th feature map activations A^k in the last conv layer

α_k^c is the scalar importance of k-th feature map obtained by averaging $\frac{\partial y^c}{\partial A^k}$ spatially

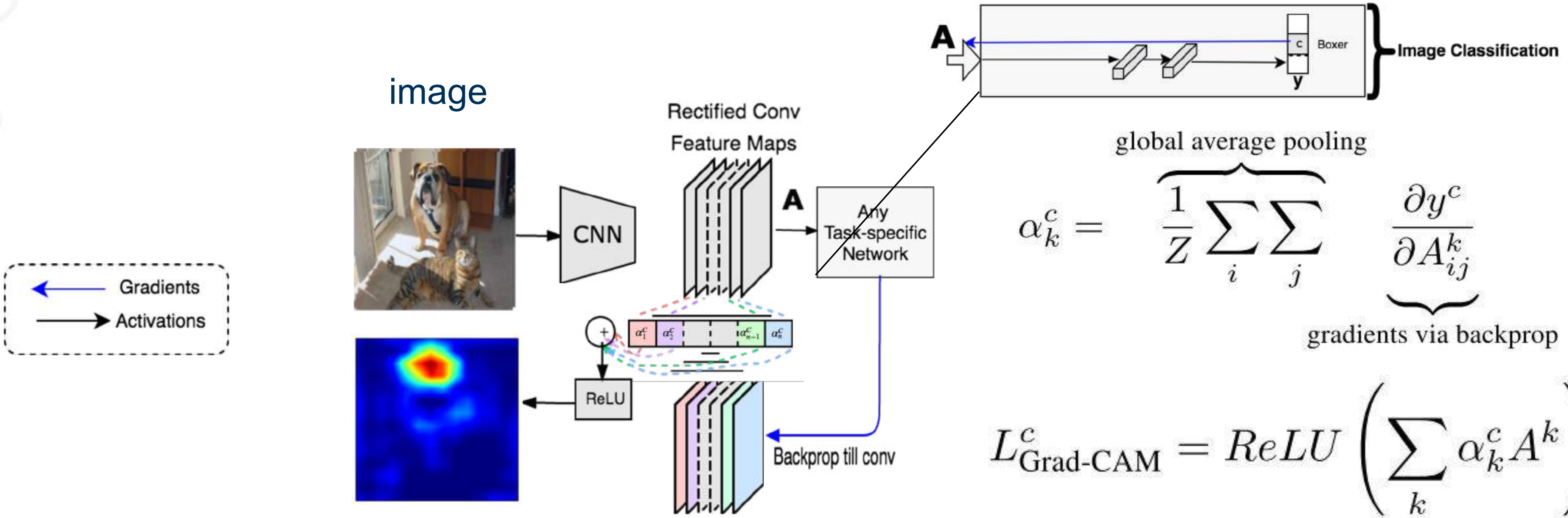
Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Gradient and Activation-based Explanations
GradCAM

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

Grad-CAM uses the gradient information flowing into the last convolutional layer of the CNN to assign importance values to each activation for a particular decision of interest.



$$\alpha_k^c = \overbrace{\frac{1}{Z}\sum_i \sum_j}^{\text{global average pooling}} \underbrace{\frac{\partial y^c}{\partial A_{ij}^k}}_{\text{gradients via backprop}}$$

$$L_{\text{Grad-CAM}}^c = ReLU \underbrace{\left(\sum_k \alpha_k^c A^k\right)}_{\text{linear combination}}$$

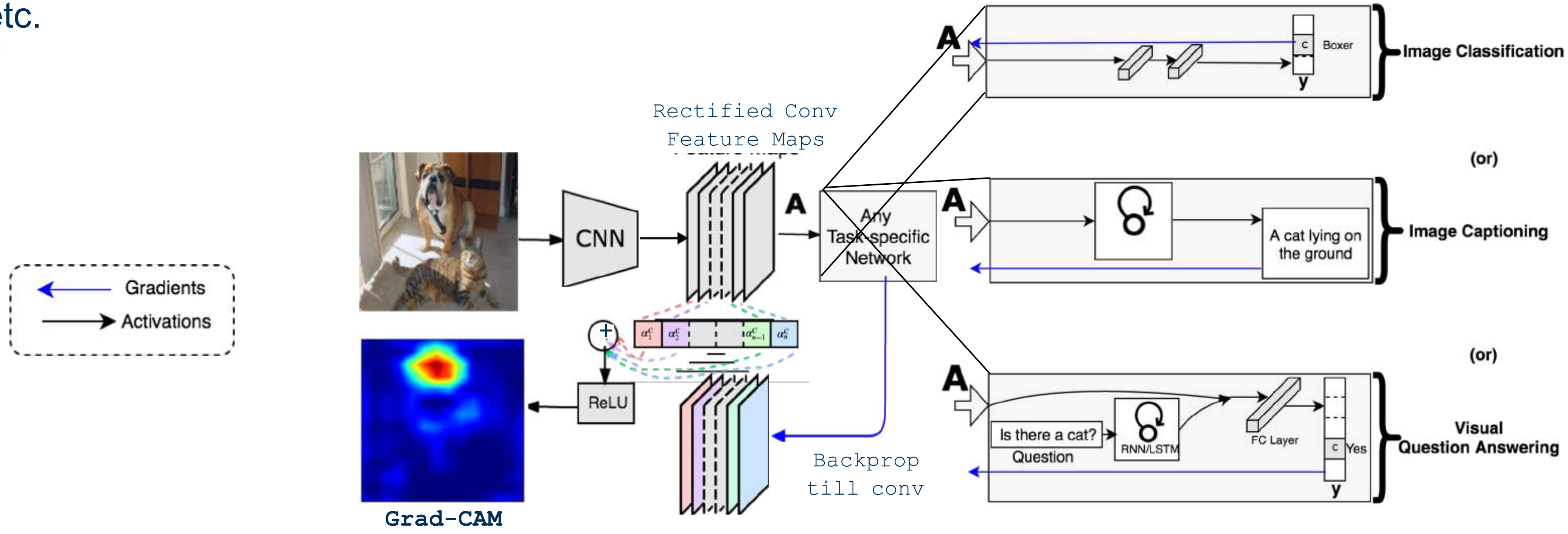
Grad-CAM (up-sampled to original image dimension)

Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Gradient and Activation-based Explanations
GradCAM

Grad-CAM generalizes to any task:

- Image classification
- Image captioning
- Visual question answering
- etc.



Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Gradient and Activation-based Explanations

Extensions of GradCAM

SCAN ME

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

GradCAM provides answers to *'Why P?'* questions. But different stakeholders require relevant and contextual explanations

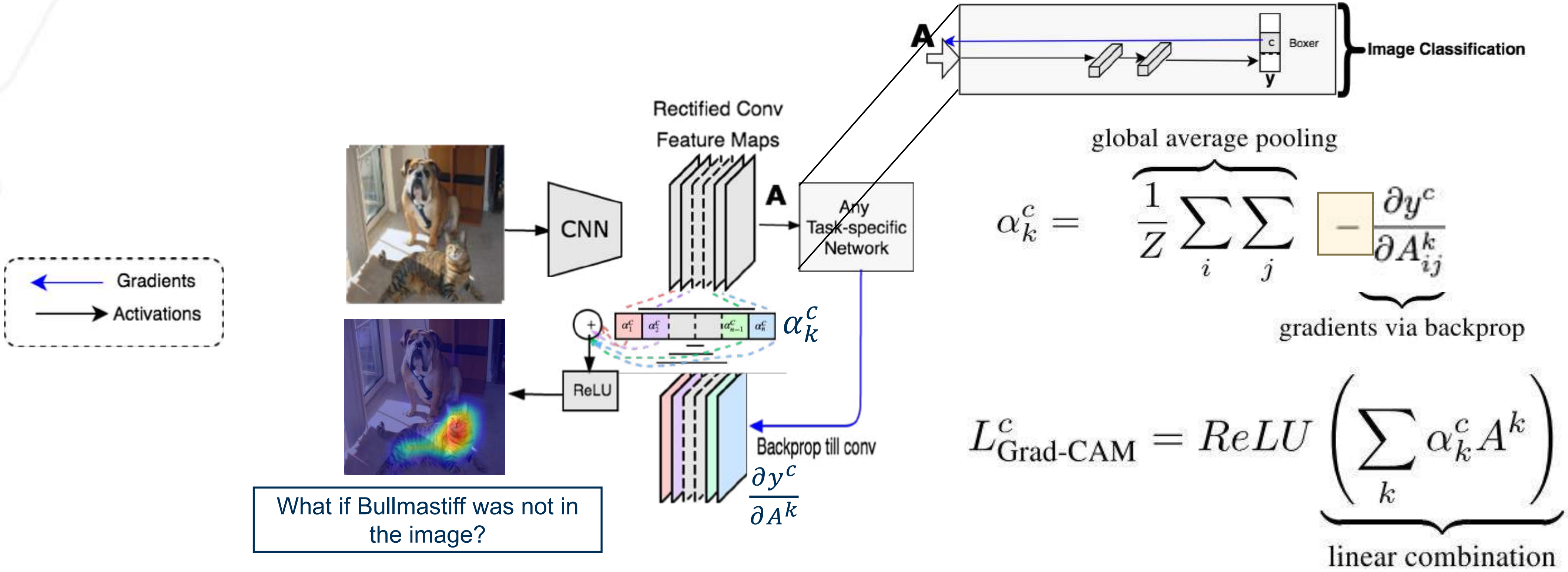
Bullmastiff	Why Bullmastiff?	What if Bullmastiff was not in the image?	Why Bullmastiff, rather than a Boxer?

AlRegib, G., & Prabhushankar, M. (2022). Explanatory Paradigms in Neural Networks: Towards relevant and contextual explanations. *IEEE Signal Processing Magazine*, *39*(4), 59-72.

Gradient and Activation-based Explanations

CounterfactualCAM: What if P is not there in the Image?

In GradCAM, global average pool the negative of gradients to obtain α^c for each kernel k



Negating the gradients effectively removes these regions from analysis

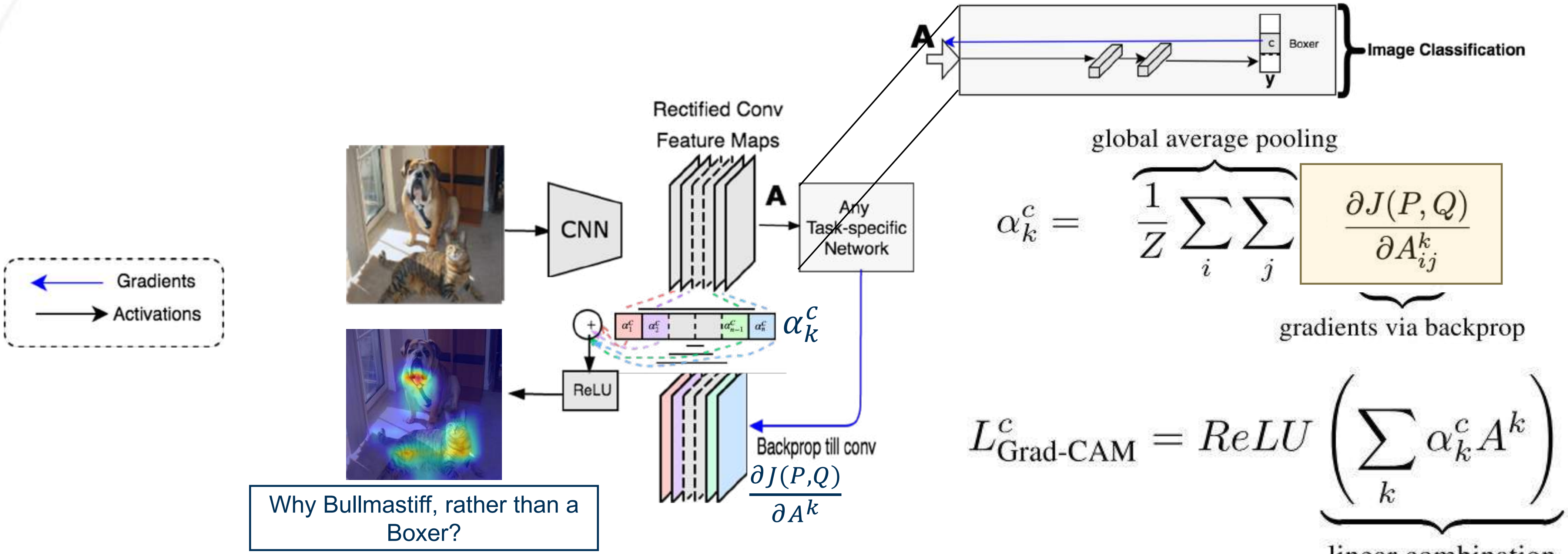
Selvaraju, Ramprasaath R., et al. "Grad-cam: Visual explanations from deep networks via gradient-based localization." *Proceedings of the IEEE international conference on computer vision*. 2017.

Gradient and Activation-based Explanations

ContrastCAM: Why P, rather than Q?

SCAN ME

In GradCAM, backward pass the loss between predicted class P and some contrast class Q to last conv layer



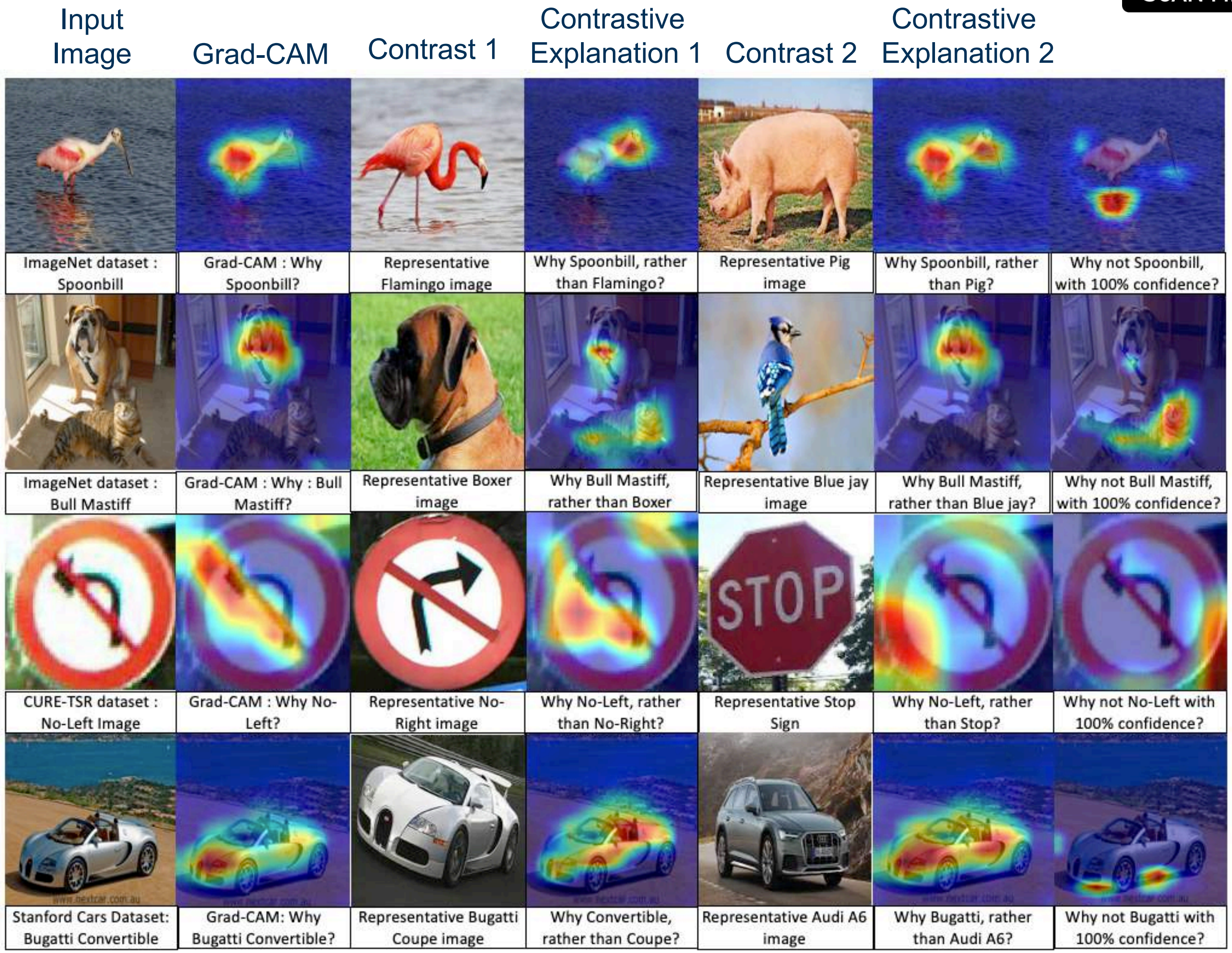
Contrast-CAM

Backpropagating the loss highlights the differences between classes P and Q.

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Gradient and Activation-based Explanations

Results of GardCAM, CounterfactualCAM, and ContrastCAM



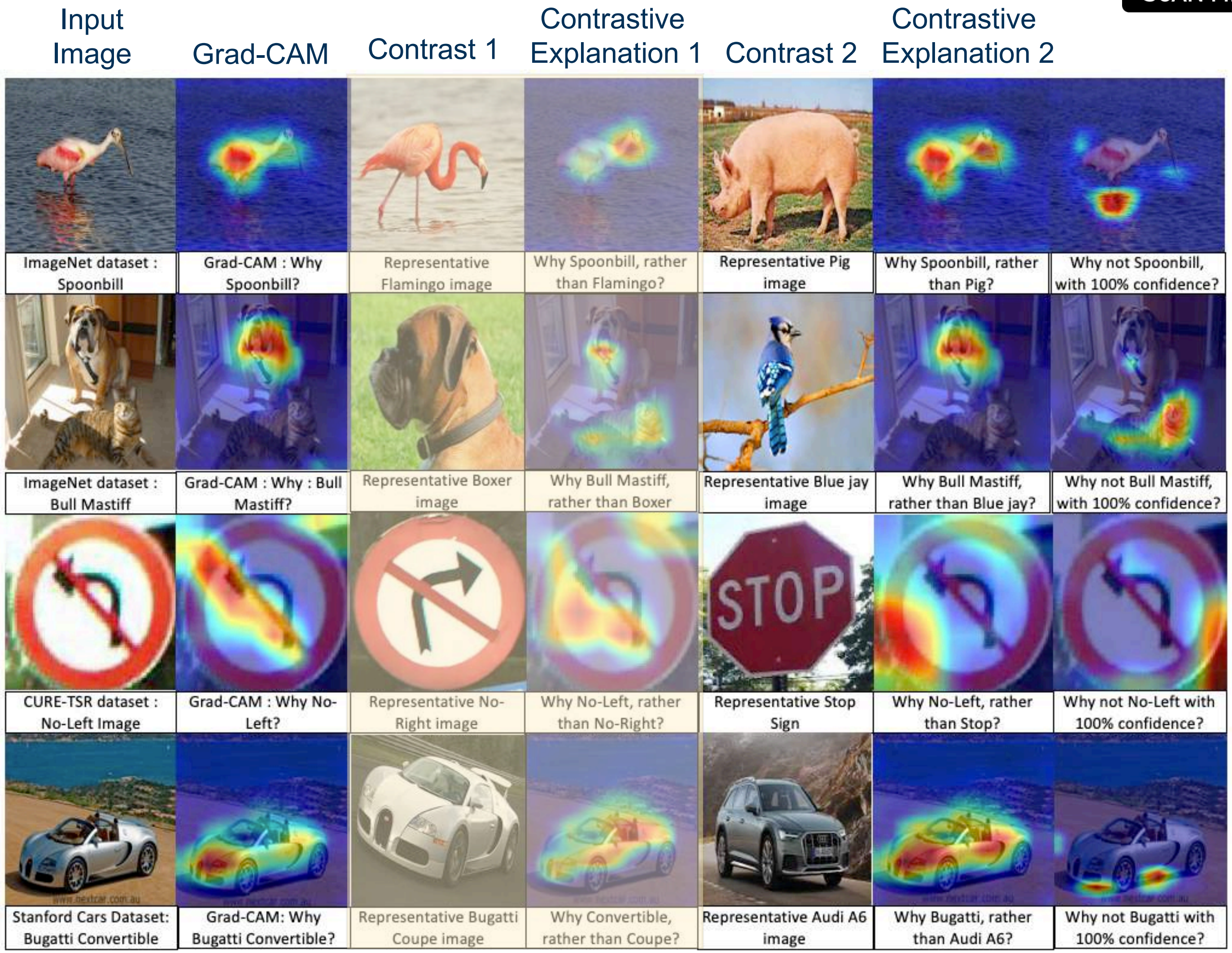
Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Gradient and Activation-based Explanations

Results of GardCAM, CounterfactualCAM, and ContrastCAM

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME



Human Interpretable

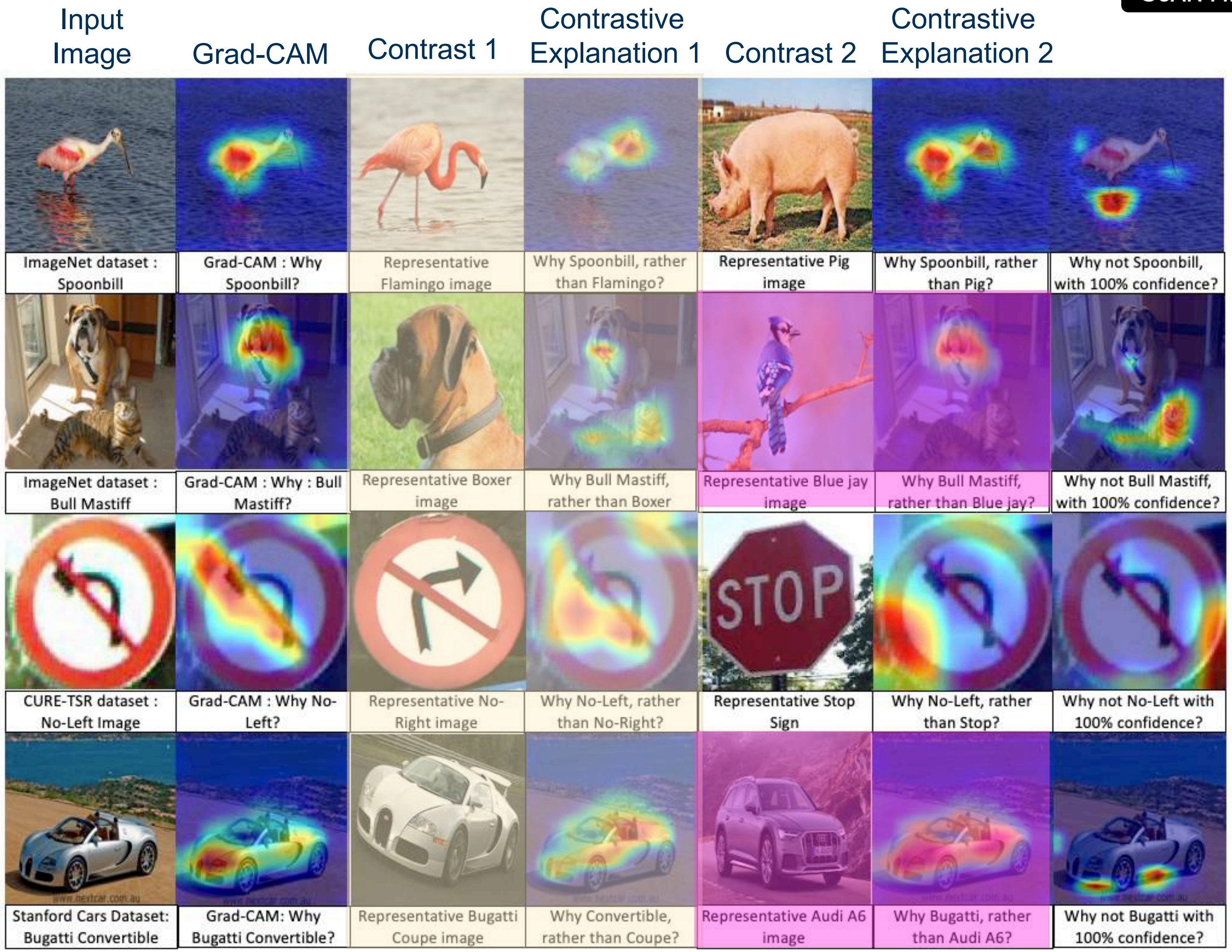
Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Gradient and Activation-based Explanations

Results of GardCAM, CounterfactualCAM, and ContrastCAM

SCAN ME

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations



Human Interpretable

Same as Grad-CAM

IEEE IV 2023

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

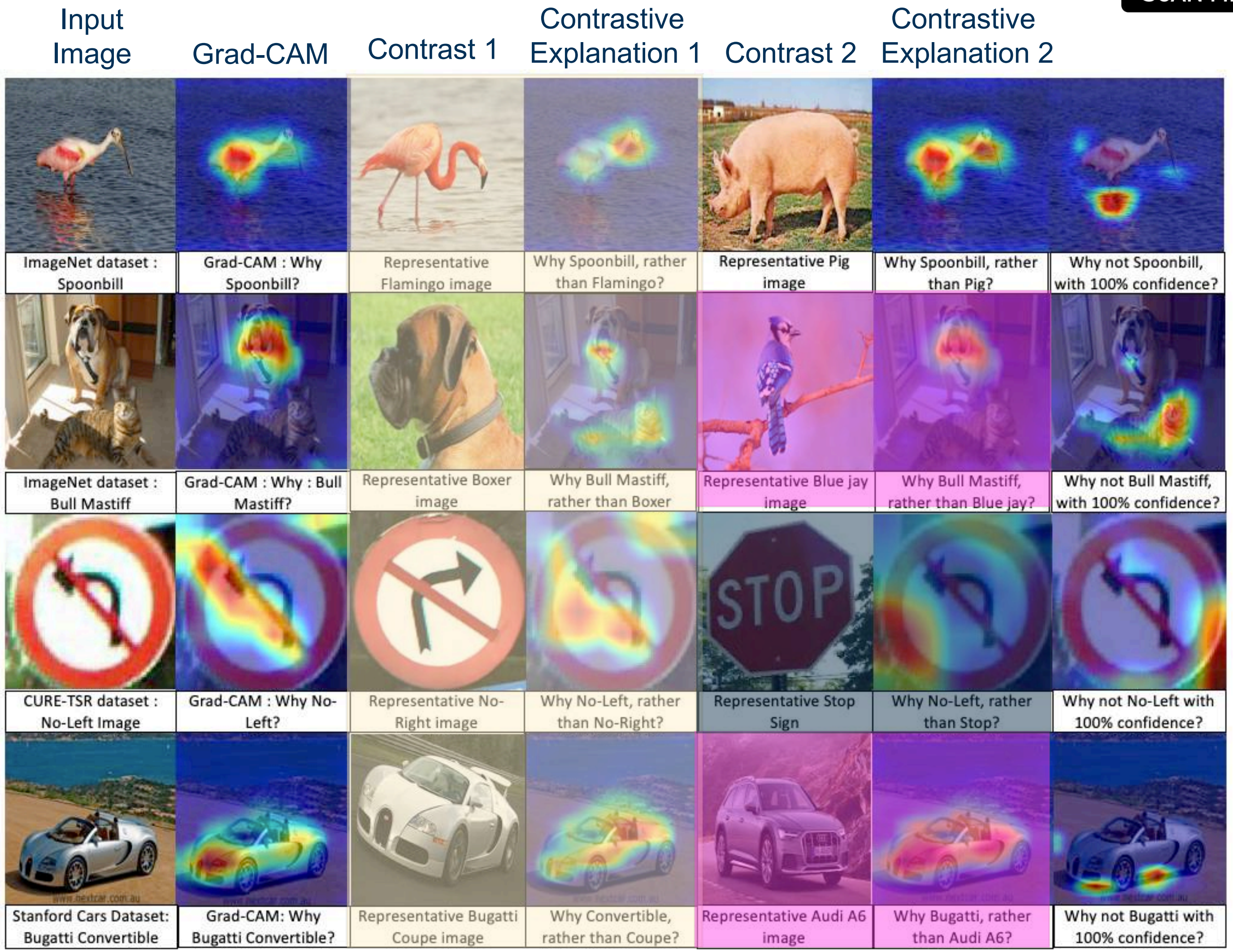
OLIVES @GeorgiaTech

Georgia Tech

SCAN ME

Gradient and Activation-based Explanations

Results of GardCAM, CounterfactualCAM, and ContrastCAM



Human Interpretable

Same as Grad-CAM

Not Human Interpretable

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Gradient and Activation-based Explanations
Results of GardCAM, CounterfactualCAM, and ContrastCAM

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

Human Interpretable

[Tutorial] | [Ghassan AlRegib and Mohit Prabhushankar] | [June 4, 2023]

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Gradient and Activation-based Explanations

Results of GardCAM, CounterfactualCAM, and ContrastCAM

Explanatory Paradigms in Neural Networks: Towards Relevant and Contextual Explanations

SCAN ME

Only traffic sign with a straight bottom-left edge – enough to say `Not STOP Sign'

Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.

Case Study 1: Leveraging anomaly scores, uncertainty scores, and explanations for Robust Recognition

NEURAL INFORMATION PROCESSING SYSTEMS

Introspective Learning: A Two-Stage Approach for Inference in Neural Networks

Mohit Prabhushankar, PhD
Postdoc

Ghassan AlRegib, PhD
Professor

SCAN ME

IEEE IV 2023 IEEE INTELLIGENT VEHICLES SYMPOSIUM

OLIVES @GeorgiaTech

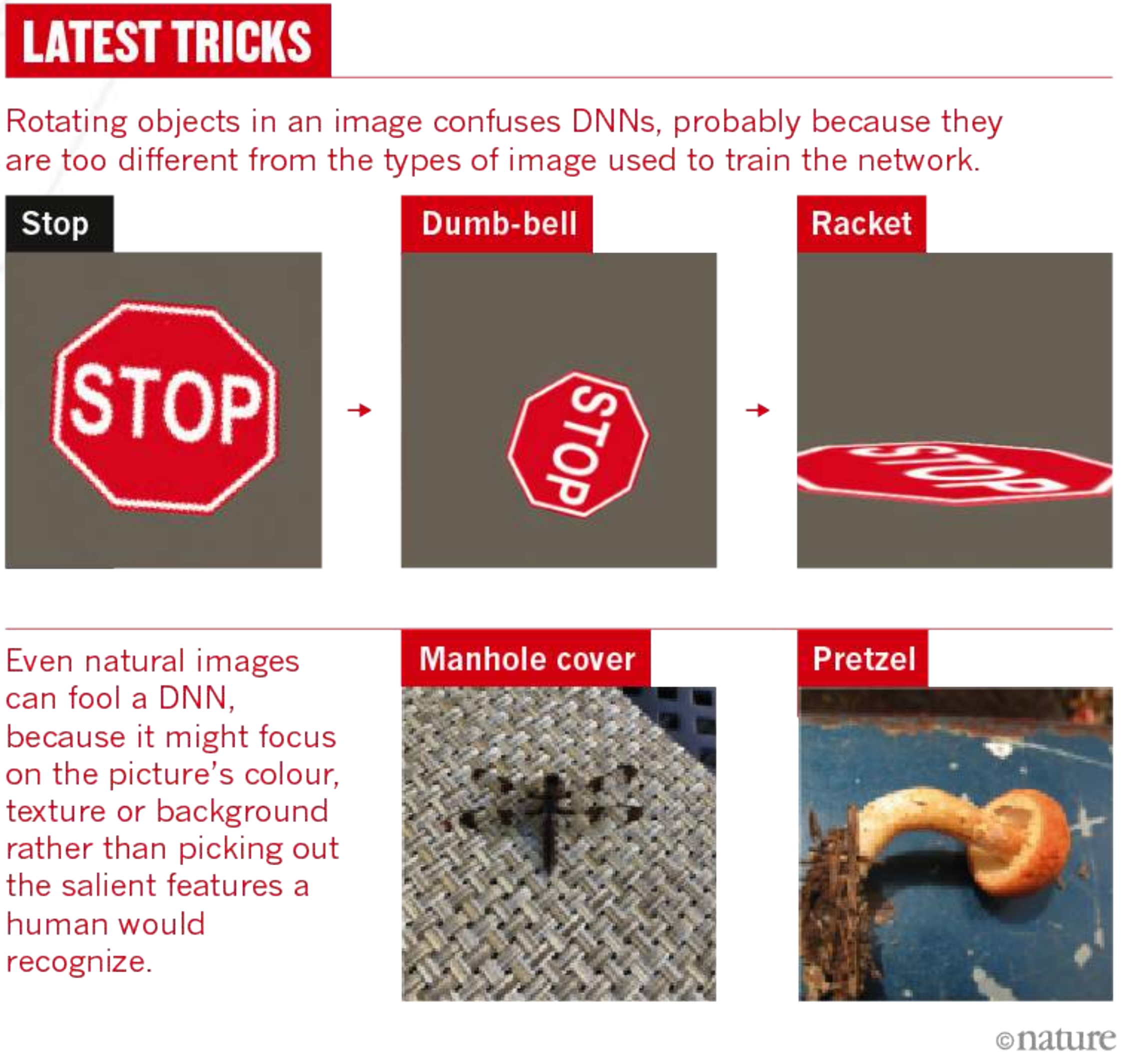
Georgia Tech

Robustness in Neural Networks

Why Robustness?

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

SCAN ME



Robustness in Neural Networks
Why Robustness?

SCAN ME

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

How would humans resolve this challenge?

We Introspect!

- Why am I being shown this slide?
- Why images of muffins rather than pastries?
- What if the dog was a bull mastiff?

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

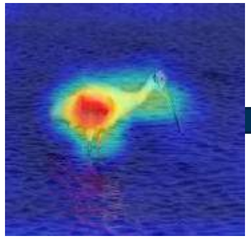
SCAN ME

Introspection Learning is a two-stage approach for Inference that combines visual sensing and reflection

Introspection

Visual Sensing

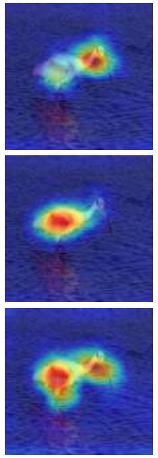
Sense pink feathers, straight beak



Spoonbill
$\hat{y}$

Feed-Forward Sensing

Reflection



Why Spoonbill, rather than Flamingo?
x does not have an S-shaped neck

Why Spoonbill, rather than Crane?
x does not have white feathers

Why Spoonbill, rather than Pig?
$x's$ leg and neck shapes are different

Spoonbill
$\tilde{y}$

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

SCAN ME

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

Introspection Learning is a two-stage approach for Inference that combines visual sensing and reflection

Goal : To simulate Introspection in Neural Networks

Definition : *We define introspections as answers to logical and targeted questions.*

What are the possible targeted questions?

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

SCAN ME

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

Introspection Learning is a two-stage approach for Inference that combines visual sensing and reflection

Bullmastiff	Why Bullmastiff?	What if Bullmastiff was not in the image?	Why Bullmastiff, rather than a Boxer?

What are the possible targeted questions?

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

SCAN ME

Introspection Learning is a two-stage approach for Inference that combines visual sensing and reflection

Goal : To simulate Introspection in Neural Networks

Contrastive Definition : *Introspection answers questions of the form `Why P, rather than Q?' where P is a network prediction and Q is the introspective class.*

Technical Definition : *Given a network* $f(x)$*, a datum* x*, and the network's prediction* $f(x) = \hat{y}$*, introspection in* $f(\cdot)$ *is the measurement of change induced in the network parameters when a label* Q *is introduced as the label for* x*..*

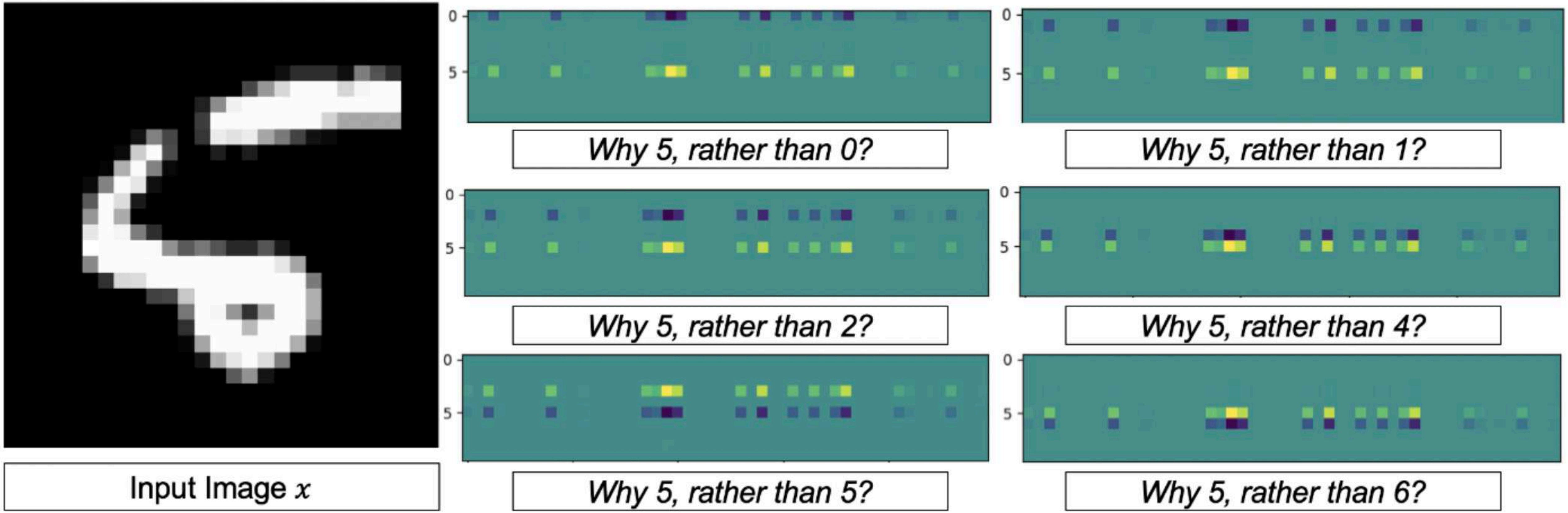
M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Introspection in Neural Networks
Gradients as Features

SCAN ME

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

For a well-trained network, the gradients are sparse and informative



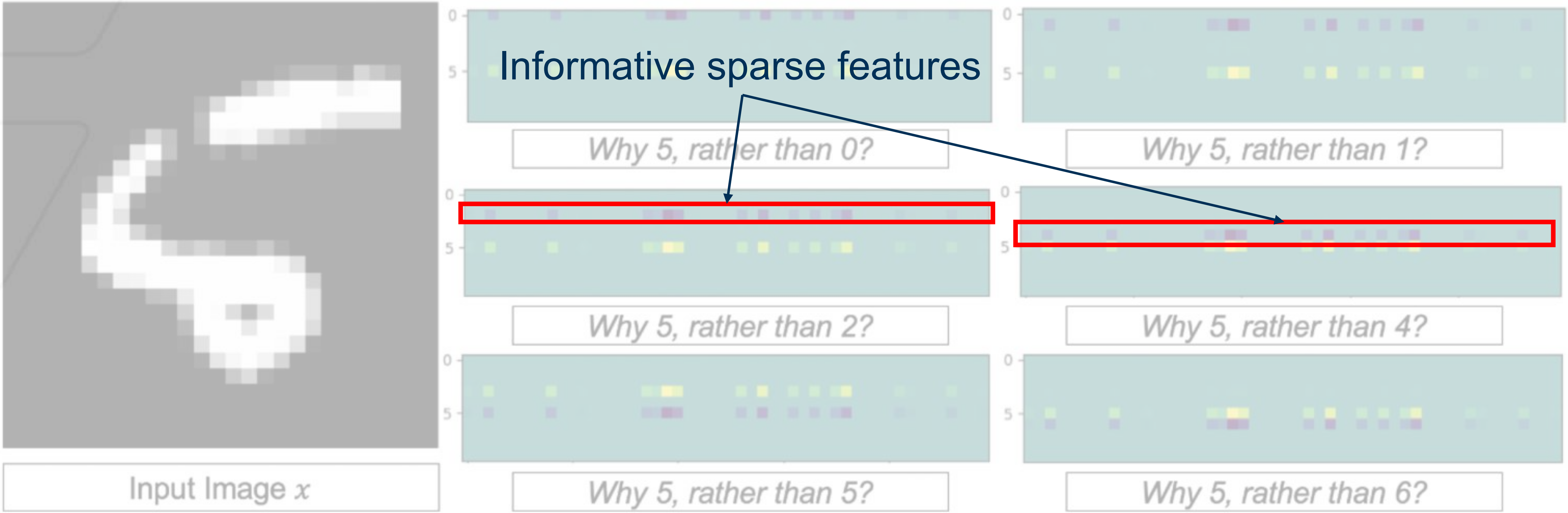
M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Introspection in Neural Networks
Gradients as Features

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

SCAN ME

For a well-trained network, the gradients are sparse and informative



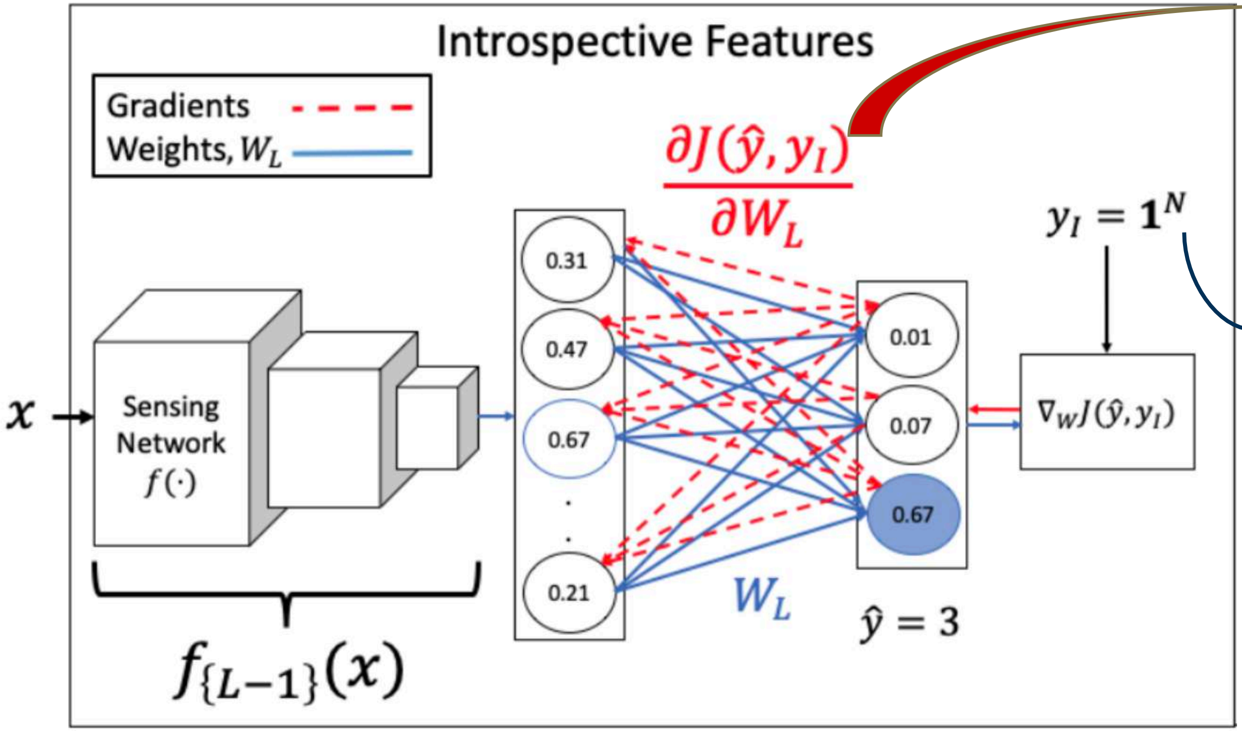
M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Introspection in Neural Networks
Deriving Gradient Features

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

SCAN ME

Measure the loss between the prediction P and a vector of all ones and backpropagate to obtain the introspective features



Normalized and vectorized gradients are introspective features

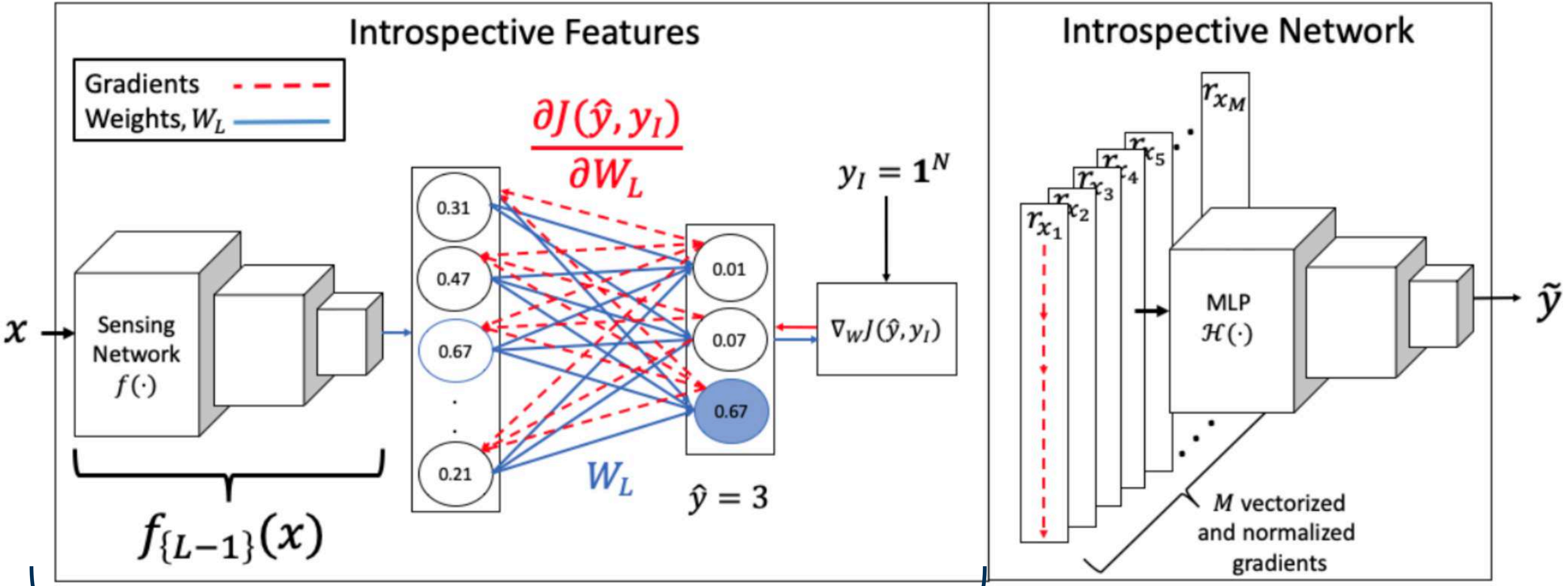
Vector of all ones: A confounding label!

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Introspection in Neural Networks
Utilizing Gradient Features

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

SCAN ME



Introspective Features

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Introspection in Neural Networks
When is Introspection Useful?

SCAN ME

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

Introspection provides robustness when the train and test distributions are different

We define robustness as being generalizable and calibrated to new testing data

Generalizable: Increased accuracy on OOD data

Calibrated: Reduces the difference between prediction accuracy and confidence

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Introspection in Neural Networks
Generalization and Calibration

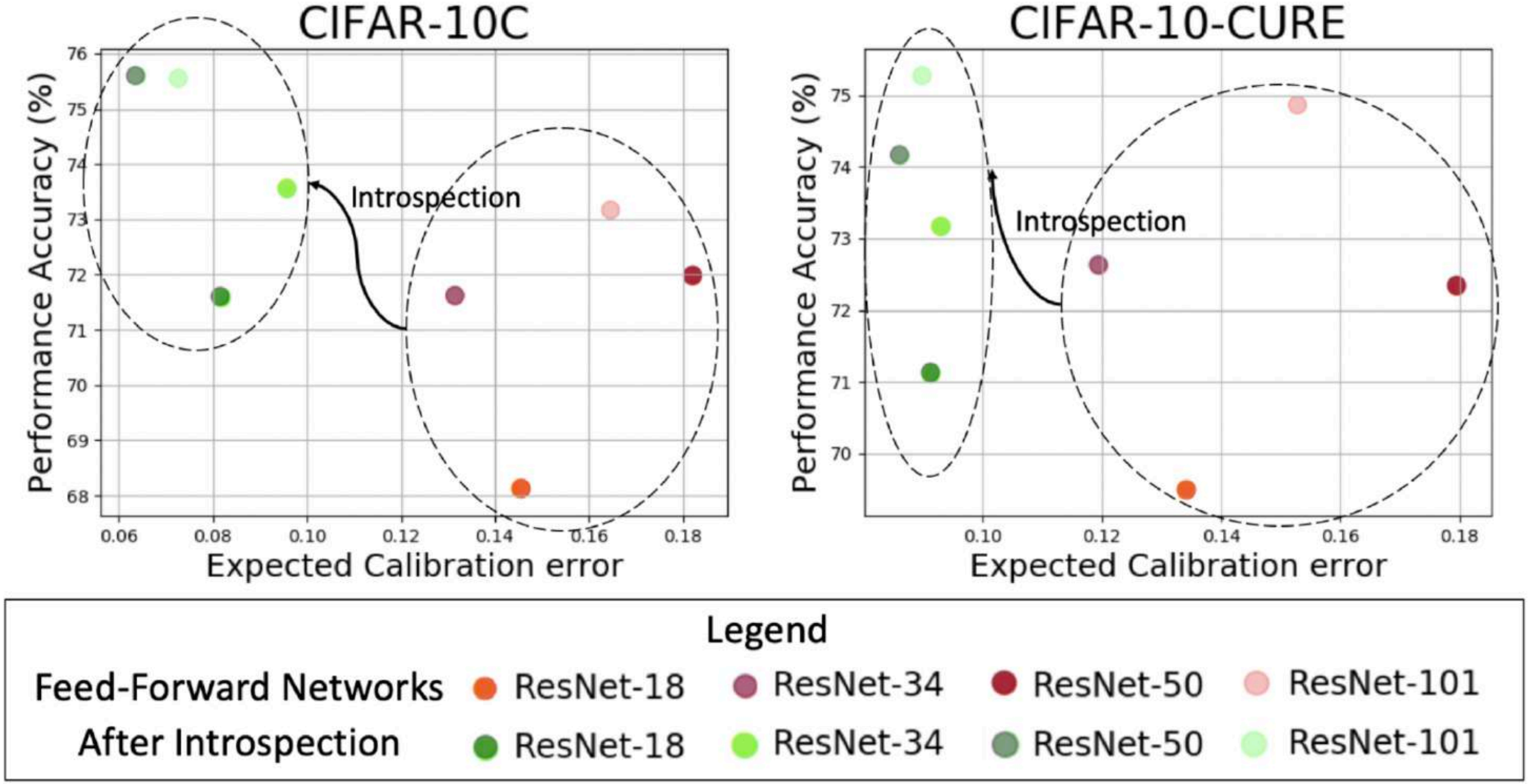
Introspective Learning: A Two-stage Approach for Inference in Neural Networks

SCAN ME

Ideal: Top-left corner

Y-Axis: Generalization

X-Axis: Calibration



M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

SCAN ME

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

Introspection is a light-weight option to resolve robustness issues

Table 1: Introspecting on top of existing robustness techniques.

METHODS		ACCURACY
RESNET-18	FEED-FORWARD	67.89%
	INTROSPECTIVE	**71.4%**
DENOISING	FEED-FORWARD	65.02%
	INTROSPECTIVE	**68.86%**
ADVERSARIAL TRAIN (27)	FEED-FORWARD	68.02%
	INTROSPECTIVE	**70.86%**
SIMCLR (19)	FEED-FORWARD	70.28%
	INTROSPECTIVE	**73.32%**
AUGMENT NOISE (28)	FEED-FORWARD	76.86%
	INTROSPECTIVE	**77.98%**
AUGMIX (26)	FEED-FORWARD	89.85%
	INTROSPECTIVE	**89.89%**

Introspection is a **plug-in approach** that works on all networks and on any downstream task!

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

SCAN ME

Introspective Learning: A Two-stage Approach for Inference in Neural Networks

Plug-in nature of Introspection benefits downstream tasks like OOD detection, Active Learning, and Image Quality Assessment!

Table 13: Performance of Contrastive Features against Feed-Forward Features and other Image Quality Estimators. Top 2 results in each row are highlighted.

Database	PSNR HA	IW SSIM	SR SIM	FSIMc	Per SIM	CSV	SUM MER	Feed-Forward UNIQUE	Introspective UNIQUE
					Outlier Ratio (OR, ↓)				
MULTI	0.013	0.013	**0.000**	0.016	0.004	**0.000**	**0.000**	**0.000**	**0.000**
TID13	**0.615**	0.701	0.632	0.728	0.655	0.687	**0.620**	0.640	**0.620**
					Root Mean Square Error (RMSE, ↓)				
MULTI	11.320	10.049	8.686	10.794	9.898	9.895	**8.212**	9.258	**7.943**
TID13	0.652	0.688	0.619	0.687	0.643	0.647	0.630	**0.615**	**0.596**
					Pearson Linear Correlation Coefficient (PLCC, ↑)				
MULTI	0.801	0.847	0.888	0.821	0.852	0.852	**0.901**	0.872	**0.908**
	-1	-1	0	-1	-1	-1	-1	-1	
TID13	0.851	0.832	0.866	0.832	0.855	0.853	0.861	**0.869**	**0.877**
	-1	-1	0	-1	-1	-1	0	0	
					Spearman's Rank Correlation Coefficient (SRCC, ↑)				
MULTI	0.715	**0.884**	0.867	0.867	0.818	0.849	**0.884**	0.867	**0.887**
	-1	0	0	0	-1	-1	0	0	
TID13	0.847	0.778	0.807	0.851	0.854	0.846	0.856	**0.860**	**0.865**
	-1	-1	-1	-1	0	-1	0	0	
					Kendall's Rank Correlation Coefficient (KRCC)				
MULTI	0.532	**0.702**	0.678	0.677	0.624	0.655	0.698	0.679	**0.702**
	-1	0	0	0	-1	0	0	0	
TID13	0.666	0.598	0.641	0.667	**0.678**	0.654	0.667	0.667	**0.677**
	0	-1	-1	0	0	0	0	0	

Table 2: Recognition accuracy of Active Learning strategies.

Methods	Architecture	Original Testset R-18	Original Testset R-34	Gaussian Noise R-18	Gaussian Noise R-34
Entropy (31)	Feed-Forward	0.365	0.358	0.244	0.249
	Introspective	0.365	0.359	**0.258**	**0.255**
Least (31)	Feed-Forward	0.371	0.359	0.252	0.25
	Introspective	0.373	0.362	**0.264**	**0.26**
Margin (32)	Feed-Forward	0.38	0.369	0.251	0.253
	Introspective	0.381	0.373	**0.265**	**0.263**
BALD (34)	Feed-Forward	0.393	0.368	0.26	0.253
	Introspective	0.396	0.375	**0.273**	**0.263**
BADGE (33)	Feed-Forward	0.388	0.37	0.25	0.247
	Introspective	0.39	0.37	**0.265**	**0.260**

Table 3: Out-of-distribution Detection of existing techniques compared between feed-forward and introspective networks.

Methods	OOD Datasets	FPR (95% at TPR) ↓	Detection Error ↓	AUROC ↑
		Feed-Forward/Introspective		
MSP (35)	Textures	58.74/**19.66**	18.04/**7.49**	88.56/**97.79**
	SVHN	61.41/**51.27**	16.92/**15.67**	89.39/**91.2**
	Places365	58.04/**54.43**	17.01/**15.07**	89.39/**91.3**
	LSUN-C	**27.95**/27.5	**9.42**/10.29	**96.07**/95.73
ODIN (36)	Textures	52.3/**9.31**	22.17/**6.12**	84.91/**91.9**
	SVHN	66.81/**48.52**	23.51/**15.86**	83.52/**91.07**
	Places365	**42.21**/51.87	16.23/**15.71**	**91.06**/90.95
	LSUN-C	**6.59**/23.66	**5.54**/10.2	**98.74**/ 95.87

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Case Study 2: Leveraging anomaly scores, uncertainty scores, and explanations for Anomalous object classification

Detecting and Classifying Anomalies in Artificial Intelligence Systems

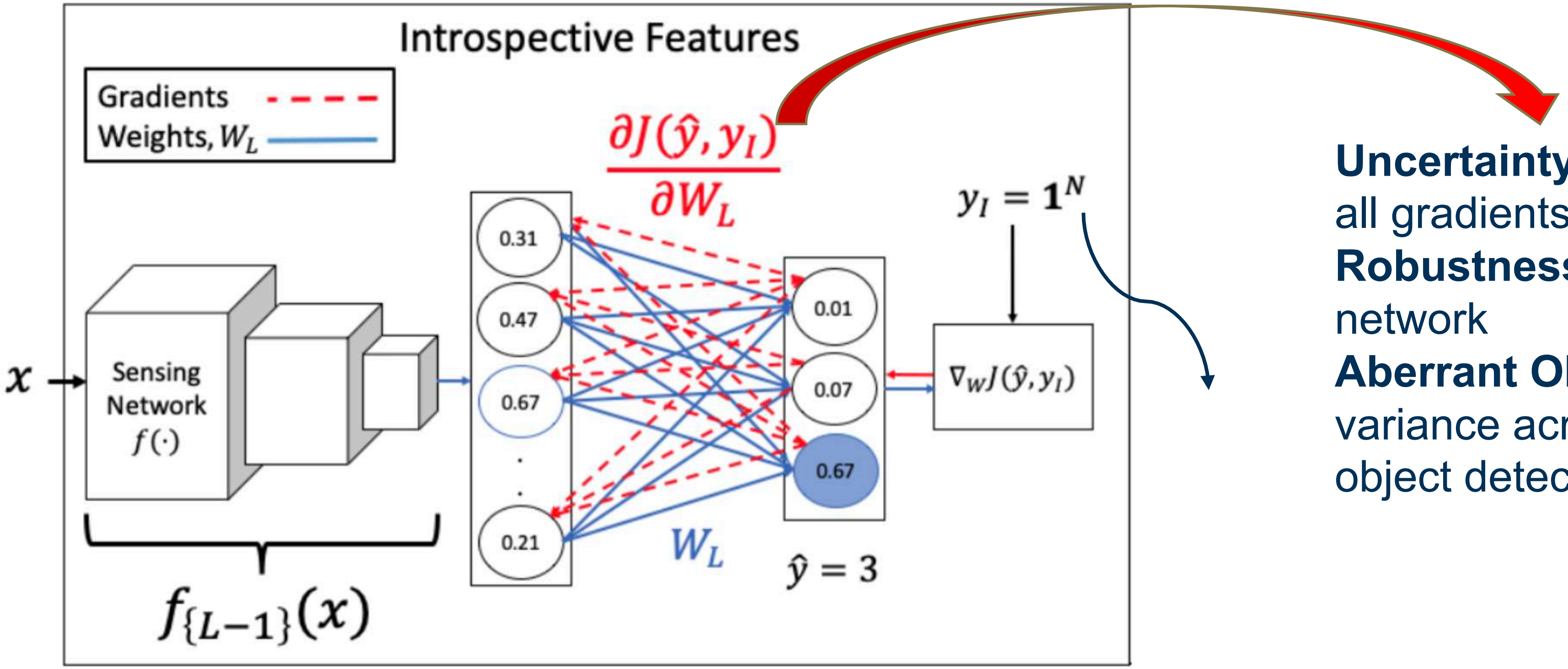
Gukyeong Kwon, PhD
Amazon AWS

Mohit Prabhushankar, PhD
Postdoc, Georgia Tech

Ghassan AlRegib, PhD
Professor, Georgia Tech

Aberrant Object Detection
Deriving Gradient Features

Measure the loss between the prediction P and a vector of all ones and backpropagate to obtain the introspective features



Uncertainty: We took energy of all gradients
Robustness: We trained a new network
Aberrant Objects: We take variance across gradients from object detector

M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022.

Aberrant Object Detection
Aberrance Detection

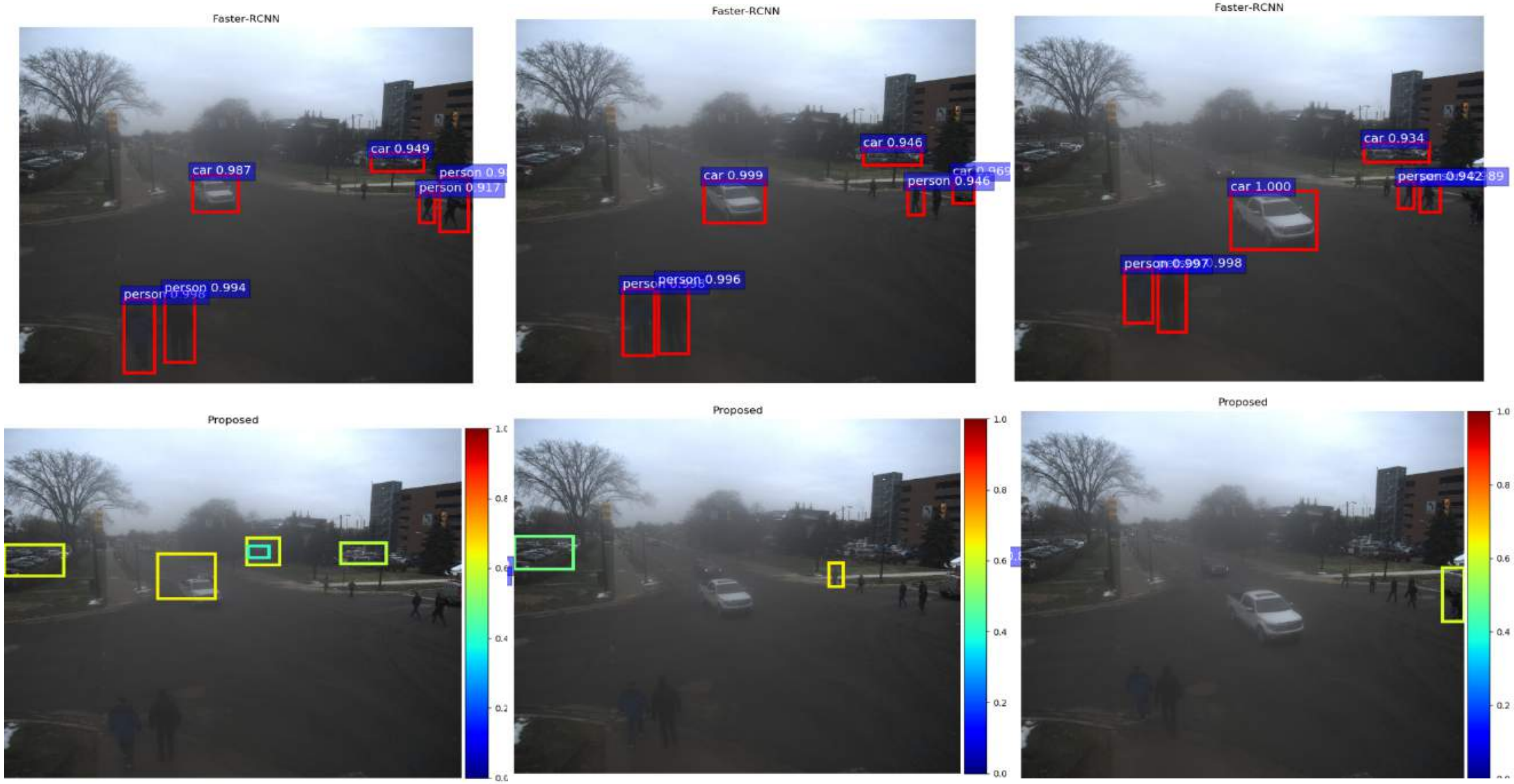
Uncertainty using variance of introspective gradients rather than energy of gradients

- Object detection algorithms would pick up on all the trained objects
- The gradient-based uncertainty approach picks up only the *aberrant* object – objects that bear a resemblance to novel classes

AlRegib, Ghassan, et al. "Detecting and Classifying Anomalies in Artificial Intelligence Systems." U.S. Patent Application No. 17/633,878.

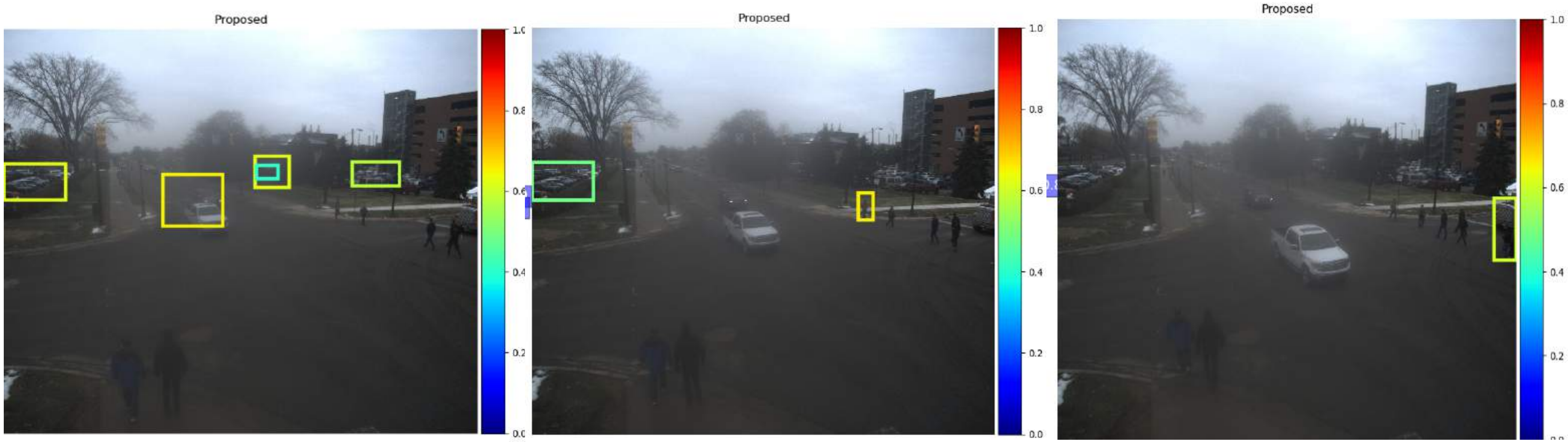
Aberrant Object Detection
Complementary to object detectors

Uncertainty using variance of introspective gradients rather than energy of gradients



AlRegib, Ghassan, et al. "Detecting and Classifying Anomalies in Artificial Intelligence Systems." U.S. Patent Application No. 17/633,878.

Use the uncertain boxes for obtaining labels from annotators



Use new annotations for subsequent training in an active learning setting

Parchami, Armin, et al. "Variance of gradient based active learning framework for training perception algorithms." U.S. Patent Application No. 17/172,854.

Objectives

Takeaways from Part III

- Part I: Challenges in Perception and Autonomy
- Part II: Deep Learning for Perception
- **Part III: Existing Deep Learning solutions to Challenges in Perception**
  - It is not always clear if aberrant events and challenges must be incorporated in training
  - Instead, they can and should be equipped with diagnostic tools at predictions
  - These diagnostic tools are anomaly and uncertainty scores for decision making and contextual explainability for post-hoc stakeholders
  - Gradients provide the change induced by an aberrant event in the network and can be used to obtain the required prediction diagnosis
- Part IV: Key Takeaways and Future Directions

A Holistic View of Perception in Intel. Vehicles

Part IV: Key Takeaways and Future Directions

Objectives

Objectives in Part IV

- Takeaway Messages and Key Insights
- Unaddressed Challenges in Perception
  - Context Awareness
  - Embedded Perception
  - V2X Perception
- Future Research Directions
  - Temporal Processing
  - Sensor Processing Architectures
  - Sensors research
  - Infrastructure + AV Datasets

Objectives

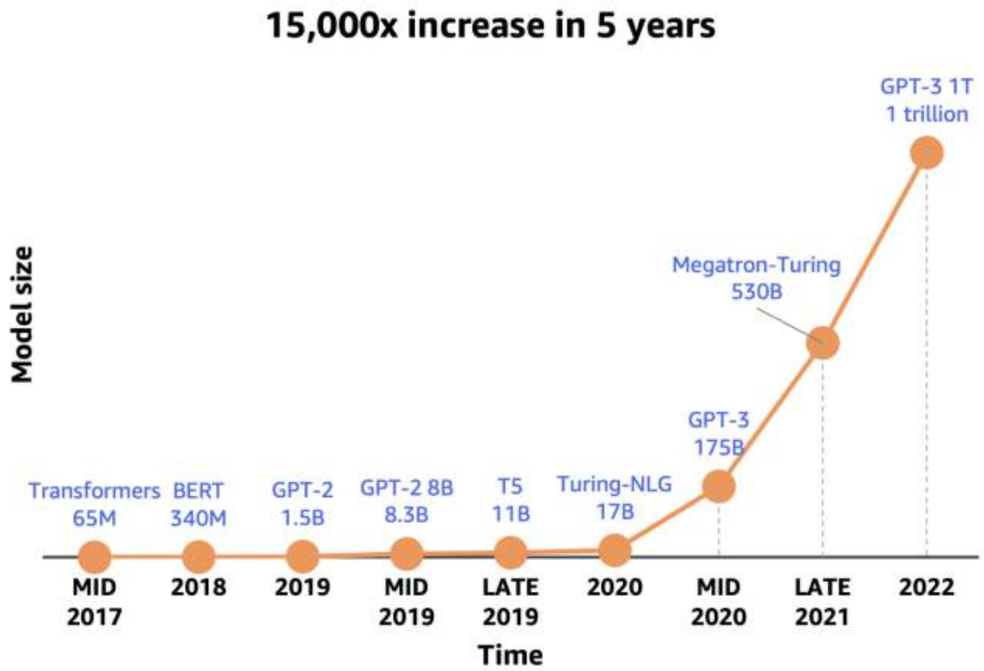
Takeaway Messages and Key Insights

- **Robustness** under challenging conditions, environments, context and surroundings-awareness are **challenges** in AV perception
  - **Deep Learning** provides a **holistic solution** to a number of the above challenges
- **Transfer Learning** and **training at scal**e help to create foundation models
  - **Self-supervised Learning** provides a framework for large scale learning on unannotated data
- It is not always clear if aberrant events and challenges must be incorporated in training
  - Instead, **model predictions** must be equipped with **diagnostic tools** at inference
  - These diagnostic tools are **anomaly and uncertainty scores** for decision making and **contextual explainability** for post-hoc stakeholders
  - **Gradients** provide the change induced by an aberrant event in the network and can be used to obtain the required **prediction diagnosis**

Perception in AVs

Unaddressed Technical Challenges for Level 3 Automation

- Challenging weather
- Challenging sensing
- Challenging environments
- Context awareness
- **Embedded perception**
- V2X perception

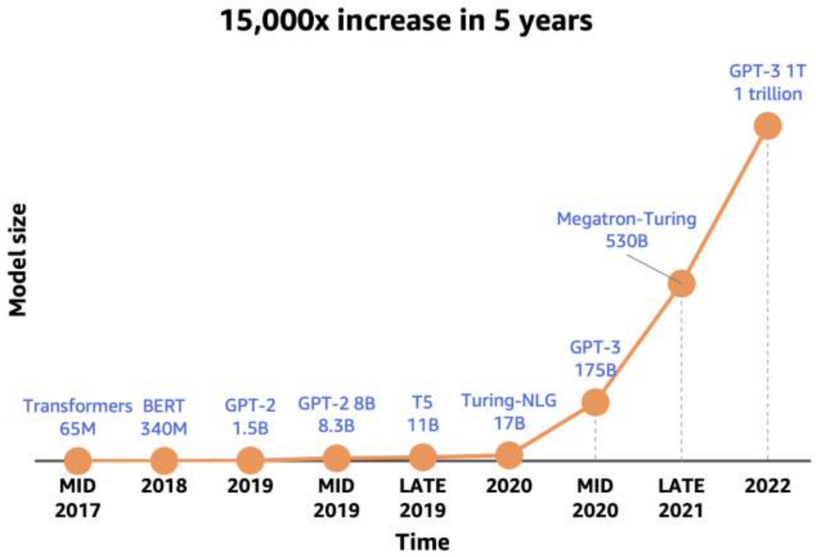


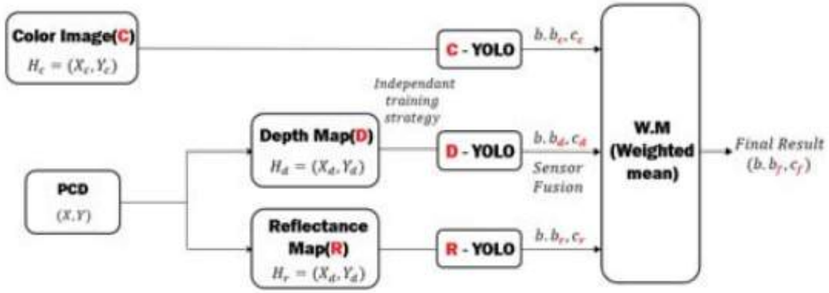
- Foundation models are great but the real-time feasibility is an issue
- The inaccuracies from model outputs is dangerous in urban settings

Perception in AVs
Unaddressed Technical Challenges for Levels 4 and 5

Foundation models with multiple sensor modalities

- Challenging weather
- Challenging sensing
- Challenging environments
- **Context awareness**
- **Embedded perception**
- **V2X perception**





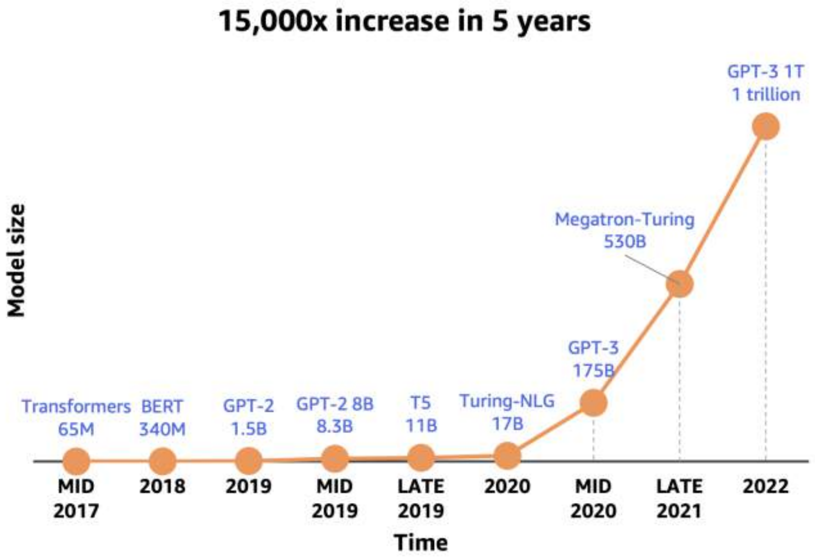
- Levels 4 and 5 automation relies on roadside infrastructure to obtain high-resolution predictions
- 10x is the rough estimate of the increase in processing power between levels of automation

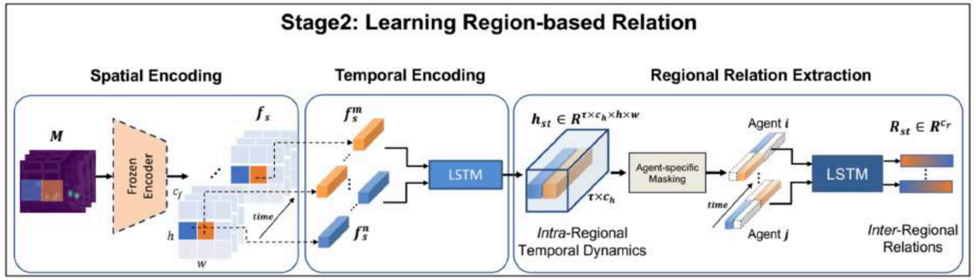
Kim, J., Kim, J., & Cho, J. (2019, December). An advanced object classification strategy using YOLO through camera and LiDAR sensor fusion. In *2019 13th International Conference on Signal Processing and Communication Systems (ICSPCS)* (pp. 1-5). IEEE.

Perception in AVs
Unaddressed Technical Challenges for Levels 4 and 5

Foundation models with multiple sensor modalities and on temporal data

- Challenging weather
- Challenging sensing
- Challenging environments
- **Context awareness**
- **Embedded perception**
- **V2X perception**



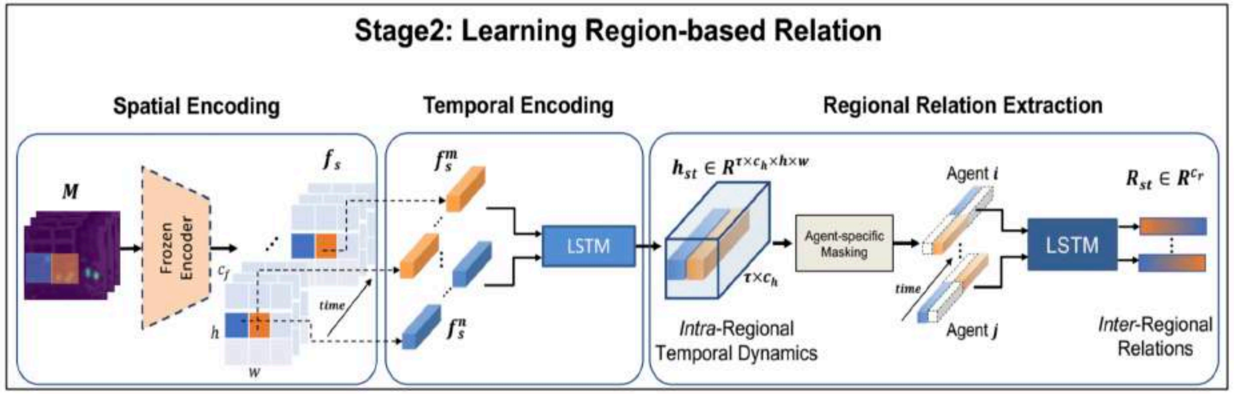


- Levels 4 and 5 automation relies on roadside infrastructure to obtain high-resolution predictions
- 10x is the rough estimate of the increase in processing power between levels of automation
- Current temporal processing = linear spatial processing in time

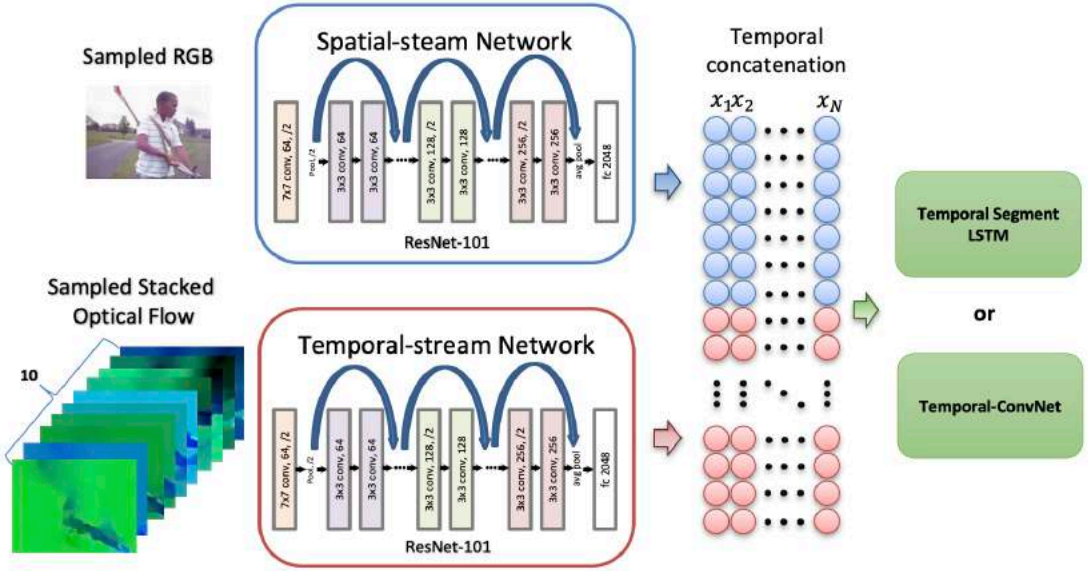
C. Zhou, G. AlRegib, A. Parchami, and K. Singh, "TrajPRed: Trajectory Prediction With Region-Based Relation Learning," *IEEE Transactions on Intelligent Transportation Systems*, submitted on Dec. 28 2022

Future Direction 1
Temporal processing of data

Temporal processing $\neq$ Linear spatial processing



Late temporal fusion: Encode all spatial data in a time-wise fashion and determine temporal relationships



Early temporal fusion: Encode both spatial and temporal information together and fuse them within the network

C. Zhou, G. AlRegib, A. Parchami, and K. Singh, "TrajPRed: Trajectory Prediction With Region-Based Relation Learning," *IEEE Transactions on Intelligent Transportation Systems*, submitted on Dec. 28 2022

Future Direction 2
Sensor processing architectures

Vision data processing was revolutionized by CNNs

Language data processing was revolutionized by Transformers

LIDAR data processing is revolutionized by ?

RADAR data processing is revolutionized by ?

…

Future Direction 3

More data with less sensors!

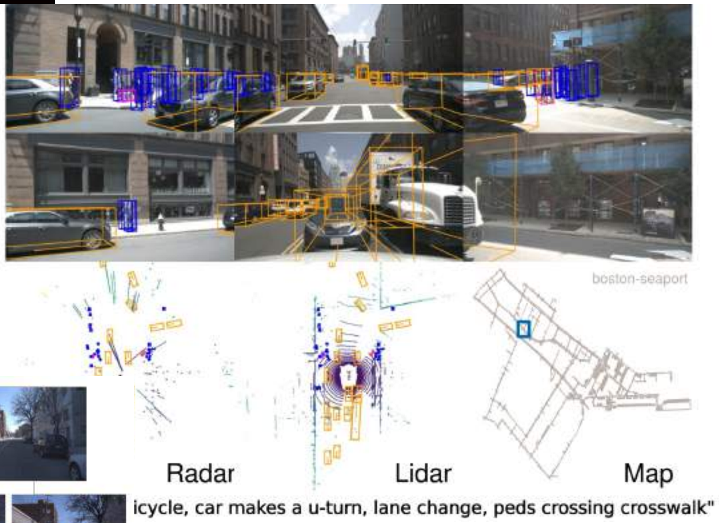
4 Fisheye cameras provide a 360 degree surround view of the car

Results from Zero-shot (i.e. using the trained model out of the box) Segment Anything Model on Woodscape dataset

Important context and objects are not segmented

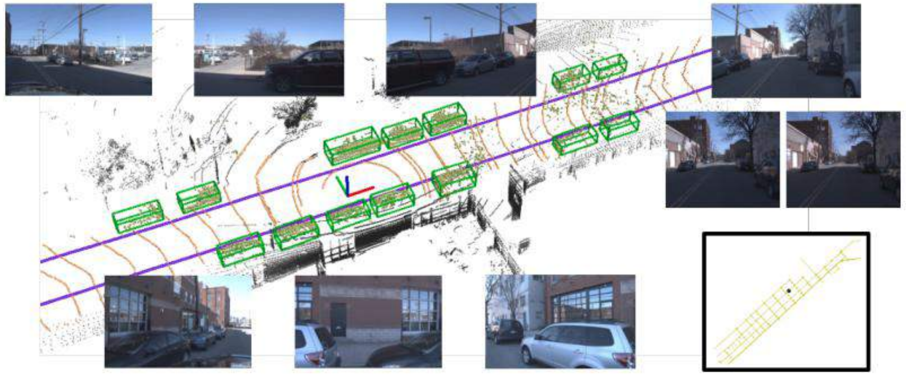
Future Direction 4
Infrastructure + AV Datasets

Abundance of egocentric AV datasets! Dearth of Infrastructure + AV datasets



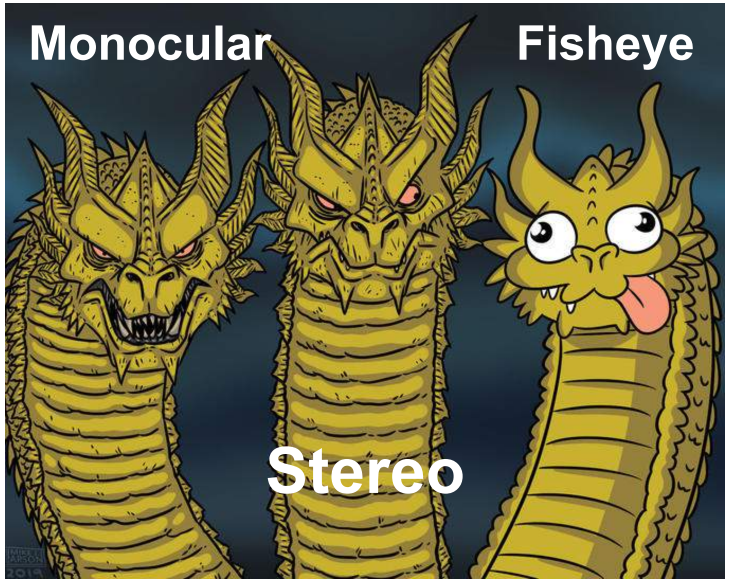
NuScenes

Argoverse



- Infrastructure datasets: Stationary sensors at traffic junctures, streets, heavy pedestrian traffic areas etc.
- Infrastructure + AV datasets: Egocentric sensors on vehicles + stationary sensors for the same scenes

Some Memes to Wrap it Up



References

Gradient representations for Robustness, OOD, Anomaly, Novelty, and Adversarial Detection

- **Gradients for robustness against noise:** M. Prabhushankar, and G. AlRegib, "Introspective Learning : A Two-Stage Approach for Inference in Neural Networks," in *Advances in Neural Information Processing Systems (NeurIPS)*, New Orleans, LA, Nov. 29 - Dec. 1 2022
- **Gradients for adversarial, OOD, corruption detection:** J. Lee, M. Prabhushankar, and G. AlRegib, "Gradient-Based Adversarial and Out-of-Distribution Detection," in *International Conference on Machine Learning (ICML) Workshop on New Frontiers in Adversarial Machine Learning*, Baltimore, MD, Jul. 2022.
- **Gradients for Open set recognition:** Lee, Jinsol, and Ghassan AlRegib. "Open-Set Recognition With Gradient-Based Representations." *2021 IEEE International Conference on Image Processing (ICIP)*. IEEE, 2021.
- **GradCon for Anomaly Detection:** Kwon, G., Prabhushankar, M., Temel, D., & AlRegib, G. (2020, August). Backpropagated gradient representations for anomaly detection. In *European Conference on Computer Vision* (pp. 206-226). Springer, Cham.
- **Gradients for adversarial, OOD, corruption detection :** J. Lee, C. Lehman, M. Prabhushankar, and G. AlRegib, "Probing the Purview of Neural Networks via Gradient Analysis," in IEEE Access, Mar. 21 2023.
- **Gradients for Novelty Detection:** Kwon, G., Prabhushankar, M., Temel, D., & AlRegib, G. (2020, October). Novelty detection through model-based characterization of neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3179-3183). IEEE.
- **Gradient-based Image Quality Assessment:** G. Kwon*, M. Prabhushankar*, D. Temel, and G. AlRegib, "Distorted Representation Space Characterization Through Backpropagated Gradients," in *IEEE International Conference on Image Processing (ICIP)*, Taipei, Taiwan, Sep. 2019.

Explainability in Neural Networks

- **Explanatory paradigms:** AlRegib, G., & Prabhushankar, M. (2022). Explanatory Paradigms in Neural Networks: Towards relevant and contextual explanations. *IEEE Signal Processing Magazine*, *39*(4), 59-72.
- **Contrastive Explanations:** Prabhushankar, M., Kwon, G., Temel, D., & AlRegib, G. (2020, October). Contrastive explanations in neural networks. In *2020 IEEE International Conference on Image Processing (ICIP)* (pp. 3289-3293). IEEE.
- **Explainabilty in Limited Label Settings:** M. Prabhushankar, and G. AlRegib, ”Extracting Causal Visual Features for Limited Label Classification,” in IEEE International Conference on Image Processing (ICIP), Sept. 2021.
- **Explainabilty through Expectancy-Mismatch:** M. Prabhushankar and G. AlRegib, "Stochastic Surprisal: An Inferential Measurement of Free Energy in Neural Networks," in Frontiers in Neuroscience, Perception Science, Volume 17, Feb. 09 2023.

References

Self Supervised Learning

- **Weakly supervised Contrastive Learning:** K. Kokilepersaud, S. Trejo Corona, M. Prabhushankar, G. AlRegib, C. Wykoff, "Clinically Labeled Contrastive Learning for OCT Biomarker Classification," in IEEE Journal of Biomedical and Health Informatics, 2023, May. 15 2023.
- **Contrastive Learning for Fisheye Images**: K. Kokilepersaud, M. Prabhushankar, Y. Yarici, G. AlRegib, and A. Parchami, "Exploiting the Distortion-Semantic Interaction in Fisheye Data," in *Open Journal of Signals Processing*, Apr. 28 2023.
- **Contrastive Learning for Severity Detection:** K. Kokilepersaud, M. Prabhushankar, G. AlRegib, S. Trejo Corona, C. Wykoff, "Gradient Based Labeling for Biomarker Classification in OCT," in *IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, Oct. 16-19 2022
- **Contrastive Learning for Seismic Images:** K. Kokilepersaud, M. Prabhushankar, and G. AlRegib, "Volumetric Supervised Contrastive Learning for Seismic Semantic Segmentation," in *International Meeting for Applied Geoscience & Energy (IMAGE)*, Houston, TX, , Aug. 28-Sept. 1 2022

Human Vision and Behavior Prediction

- **Pedestrian Trajectory Prediction:** C. Zhou, G. AlRegib, A. Parchami, and K. Singh, "TrajPRed: Trajectory Prediction With Region-Based Relation Learning," *IEEE Transactions on Intelligent Transportation Systems*, submitted on Dec. 28 2022.
- **Human Visual Saliency in trained Neural Nets:** Y. Sun, M. Prabhushankar, and G. AlRegib, "Implicit Saliency in Deep Neural Networks," in *IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, Oct. 2020.
- **Human Image Quality Assessment:** D. Temel, M. Prabhushankar and G. AlRegib, "UNIQUE: Unsupervised Image Quality Estimation," in IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1414-1418, Oct. 2016.

Open-source Datasets to assess Robustness

- **CURE-TSD:** D. Temel, M-H. Chen, and G. AlRegib, "Traffic Sign Detection Under Challenging Conditions: A Deeper Look Into Performance Variations and Spectral Characteristics," in *IEEE Transactions on Intelligent Transportation Systems*, Jul. 2019
- **CURE-TSR:** D. Temel, G. Kwon*, M. Prabhushankar*, and G. AlRegib, "CURE-TSR: Challenging Unreal and Real Environments for Traffic Sign Recognition," in *Advances in Neural Information Processing Systems (NIPS) Workshop on Machine Learning for Intelligent Transportation Systems*, Long Beach, CA, Dec. 2017
- **CURE-OR:** D. Temel*, J. Lee*, and G. AlRegib, "CURE-OR: Challenging Unreal and Real Environments for Object Recognition," in *IEEE International Conference on Machine Learning and Applications (ICMLA)*, Orlando, FL, Dec. 2018

References

Active Learning

- **Active Learning and Training with High Information Content:** R. Benkert, M. Prabhushankar, G. AlRegib, A. Parchami, and E. Corona, "Gaussian Switch Sampling: A Second Order Approach to Active Learning," in IEEE Transactions on Artificial Intelligence (TAI), Feb. 05 2023
- **Active Learning Dataset on vision and LIDAR data:** Y. Logan, R. Benkert, C. Zhou, K. Kokilepersaud, M. Prabhushankar, G. AlRegib, K. Singh, E. Corona and A. Parchami, "FOCAL: A Cost-Aware Video Dataset for Active Learning," IEEE Transactions on Circuits and Systems for Video Technology, submitted on Apr. 29 2023
- **Active Learning on OOD data:** R. Benkert, M. Prabhushankar, and G. AlRegib, "Forgetful Active Learning With Switch Events: Efficient Sampling for Out-of-Distribution Data," in *IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, Oct. 16-19 2022
- **Active Learning for Biomedical Images**: Y. Logan, R. Benkert, A. Mustafa, G. Kwon, G. AlRegib, "Patient Aware Active Learning for Fine-Grained OCT Classification," in *IEEE International Conference on Image Processing (ICIP)*, Bordeaux, France, Oct. 16-19 2022

Uncertainty Estimation

- **Gradient-based Uncertainty:** J. Lee and G. AlRegib, "Gradients as a Measure of Uncertainty in Neural Networks," in *IEEE International Conference on Image Processing (ICIP)*, Abu Dhabi, United Arab Emirates, Oct. 2020
- **Uncertainty Visualization in Seismic Images:** R. Benkert, M. Prabhushankar, and G. AlRegib, "Reliable Uncertainty Estimation for Seismic Interpretation With Prediction Switches," in *International Meeting for Applied Geoscience & Energy (IMAGE)*, Houston, TX, , Aug. 28-Sept. 1 2022.
- **Uncertainty and Disagreements in Label Annotations:** C. Zhou, M. Prabhushankar, and G. AlRegib, "On the Ramifications of Human Label Uncertainty," in *NeurIPS 2022 Workshop on Human in the Loop Learning*, Oct. 27 2022
- **Uncertainty in Saliency Estimation:** T. Alshawi, Z. Long, and G. AlRegib, "Unsupervised Uncertainty Estimation Using Spatiotemporal Cues in Video Saliency Detection," in *IEEE Transactions on Image Processing*, vol. 27, pp. 2818-2827, Jun. 2018.