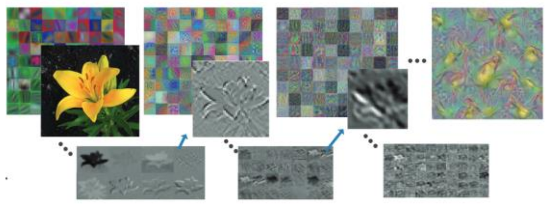
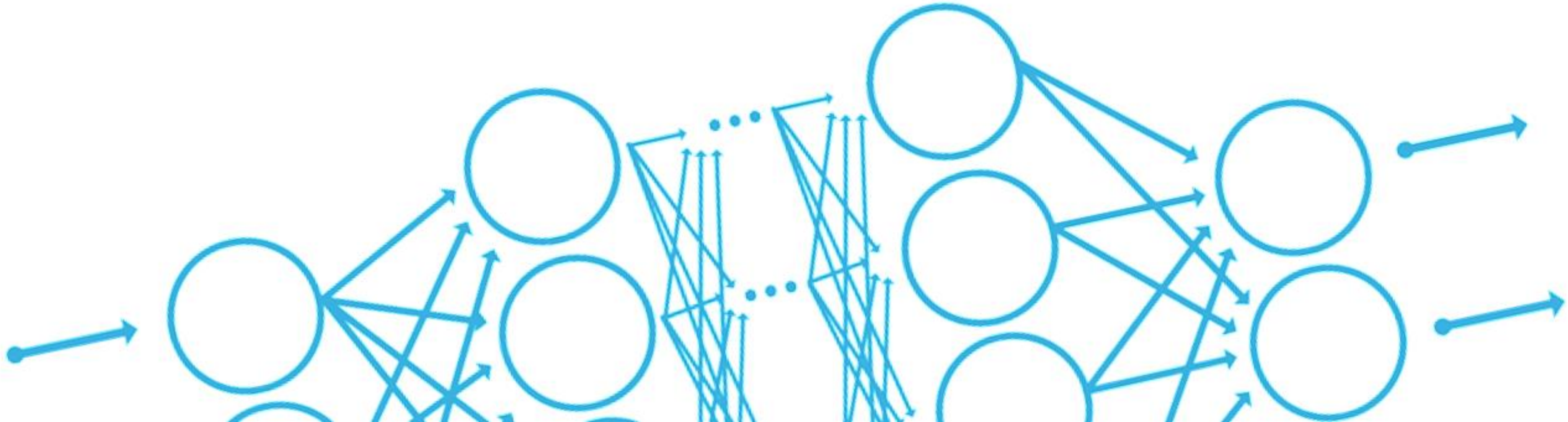
Demystifying Deep Learning

A Practical Approach in MATLAB



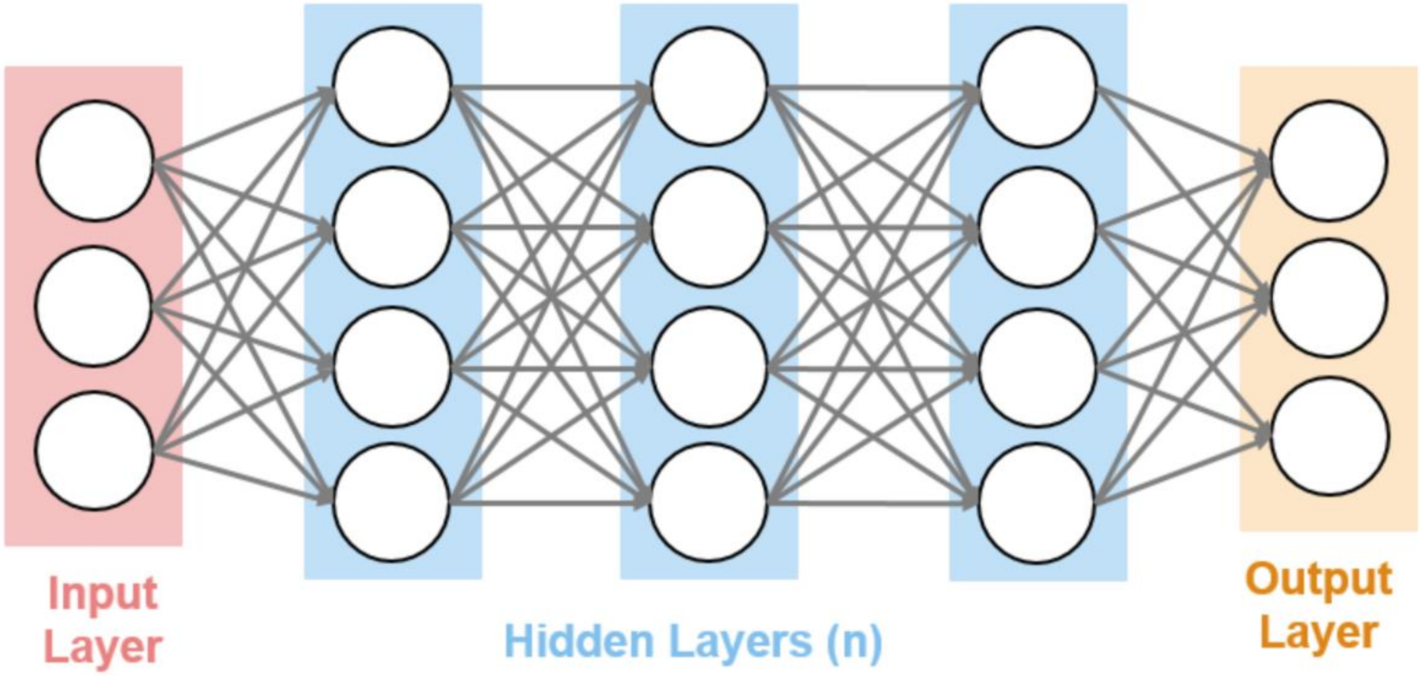
© 2018 The MathWorks, Inc.

What is Deep Learning?

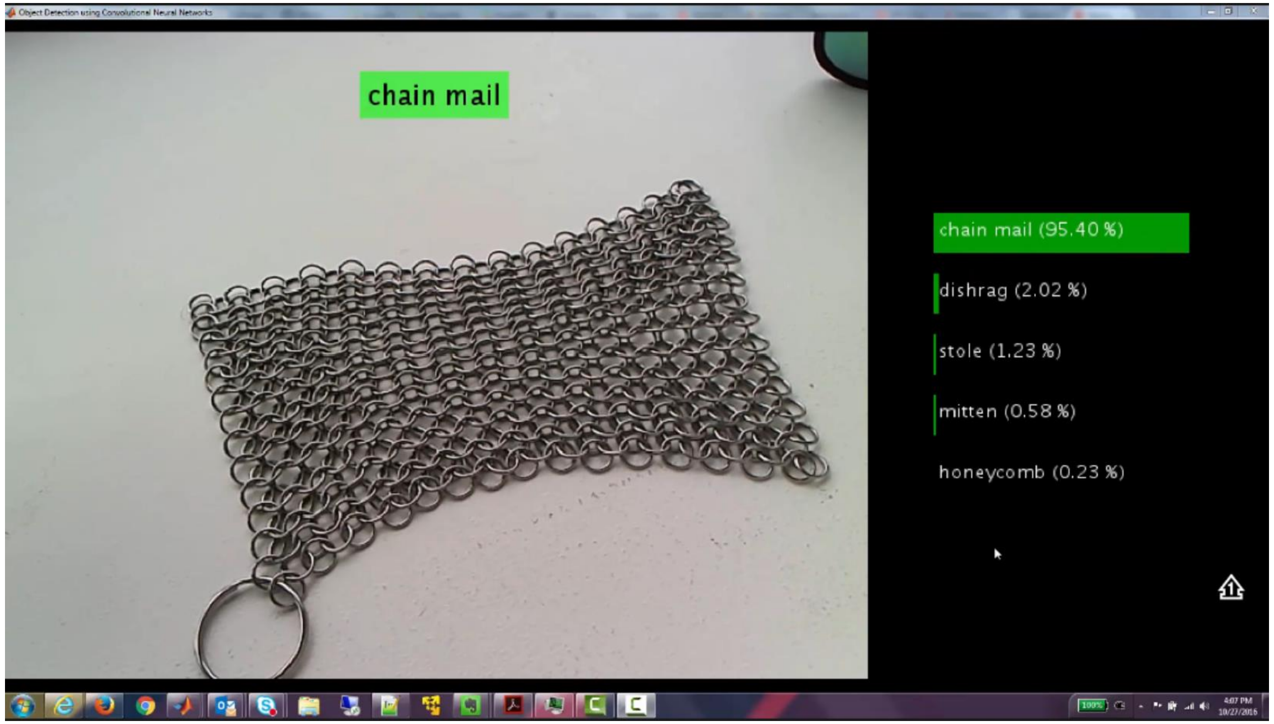


Deep learning is a type of machine learning in which a model learns to perform tasks directly from image, time-series or text data.

Deep learning is usually implemented using a **neural network architecture.**

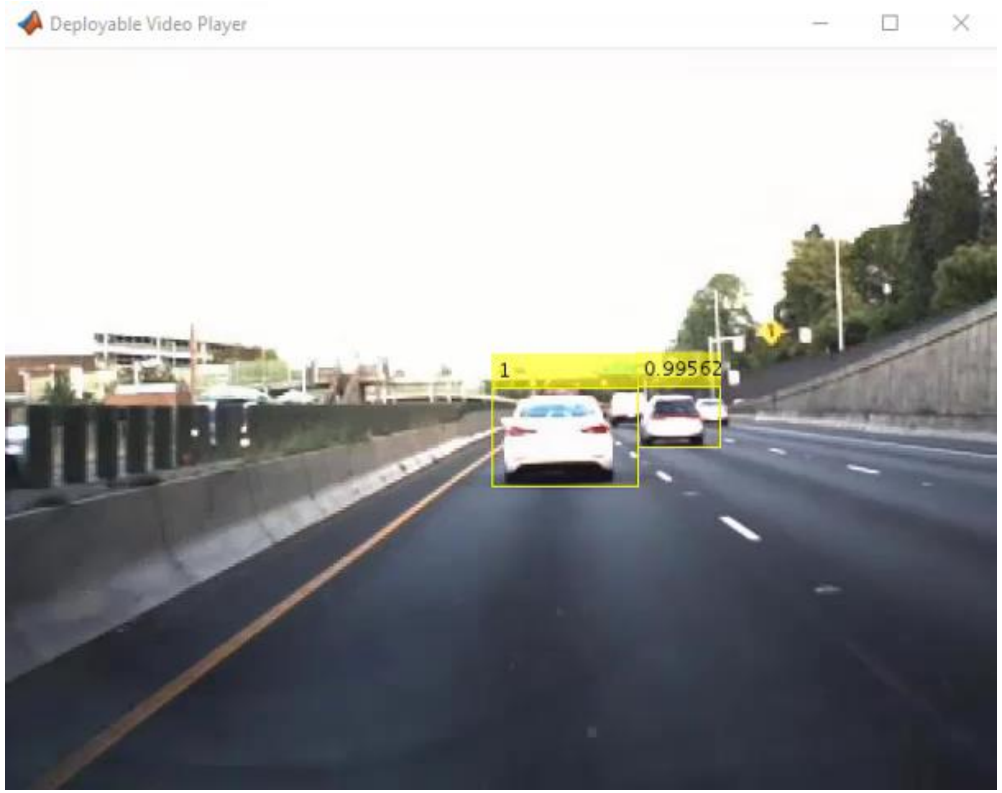


Example 1: Object recognition using deep learning



Training (GPU)	Millions of images from 1000 different categories
Prediction	Real-time object recognition using a webcam connected to a laptop

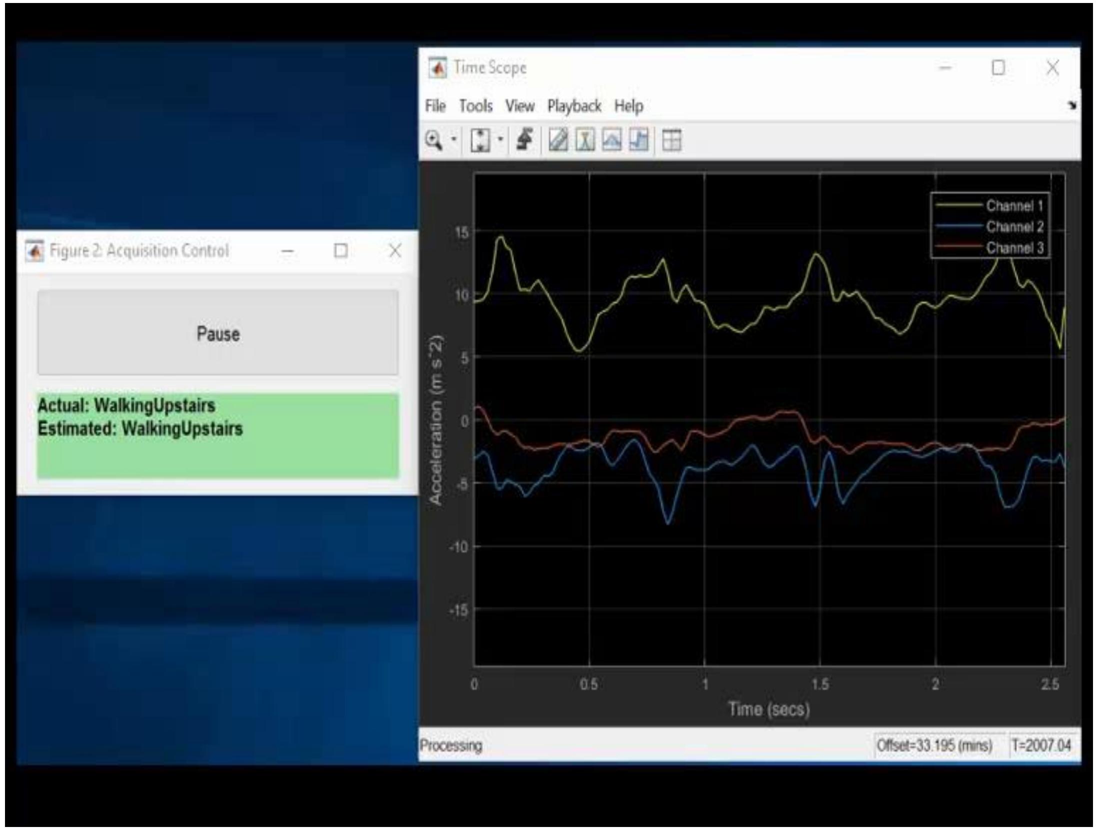
Example 2: Detection and localization using deep learning

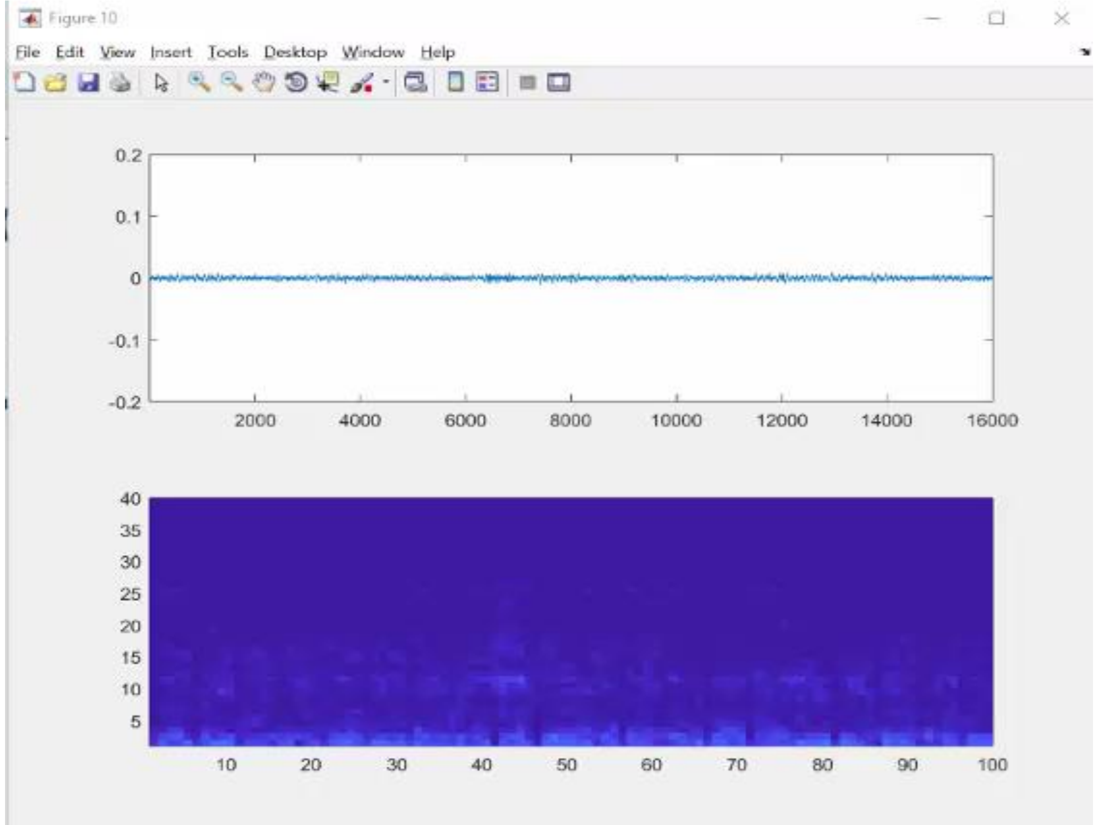


Regions with Convolutional Neural Network Features (R-CNN)

Semantic Segmentation using SegNet

Example 3: Analyzing signal data using deep learning





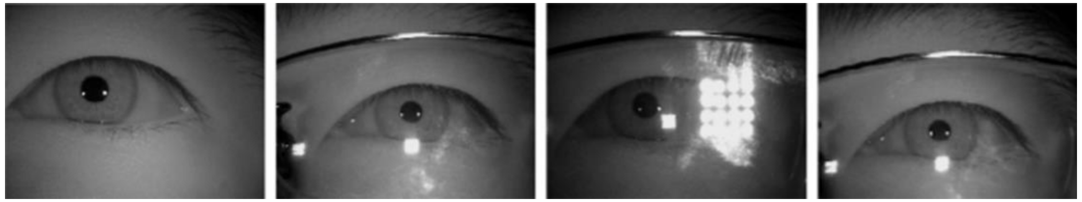
Signal Classification using LSTMs

Speech Recognition using CNNs

Agenda

- Why deep learning?
- Fashion MNIST: The “Hello, World!” of deep learning
- Transfer learning with CNNs
- (optional) Semantic segmentation
- (optional) Deep learning with time series data
- Ground Truth Labeling for datasets
- Everything else in deep learning…

Diverse Applications of Deep Learning



Iris Recognition – 99.4% accuracy[1]

MatConvnet

Rain Detection and Removal[3]

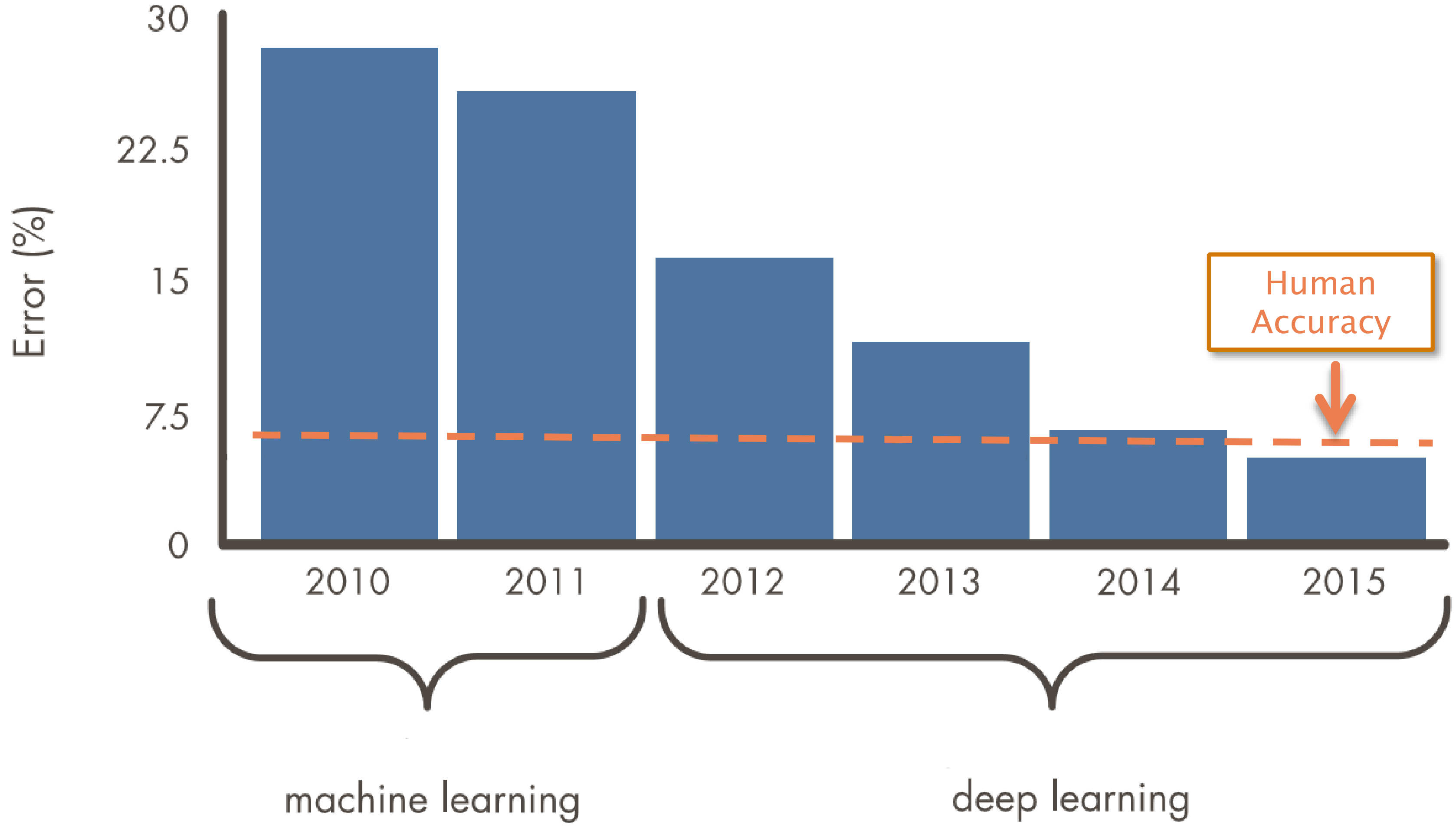
MatCaffe

Human Aware Navigation for Robots[2]

MatConvnet

1. Source: An experimental study of deep convolutional features for iris recognition Signal Processing in Medicine and Biology Symposium (SPMB), 2016 IEEE Shervin Minaee ; Amirali Abdolrashidiy ; Yao Wang; An experimental study of deep convolutional features for iris recognition
2. "A Real-Time Pedestrian Detector using Deep Learning for Human-Aware Navigation" David Ribeiro, Andre Mateus, Jacinto C. Nascimento, and Pedro Miraldo
3. Deep Joint Rain Detection and Removal from a Single Image" Wenhan Yang, Robby T. Tan, Jiashi Feng, Jiaying Liu, Zongming Guo, and Shuicheng Yan

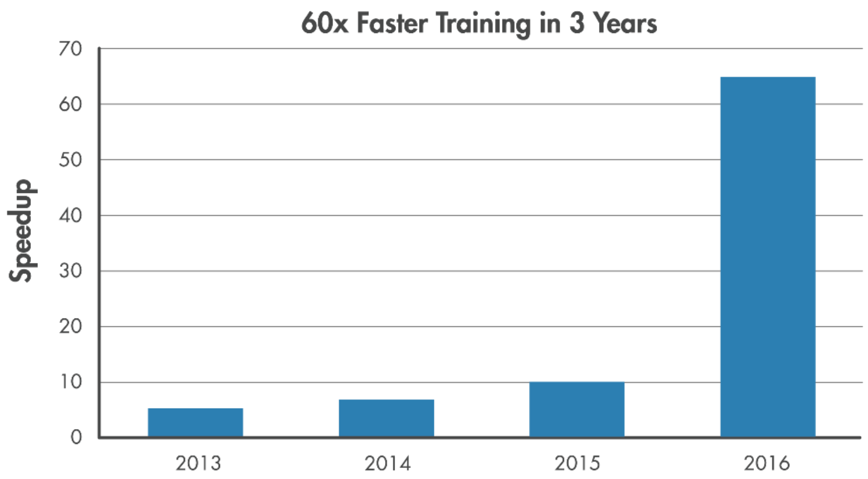
Why is Deep Learning So Popular Now?



Source: ILSVRC Top-5 Error on ImageNet

Deep Learning Enablers

- Increased GPU acceleration



- World-class models

AlexNet
PRETRAINED MODEL

VGG-16
PRETRAINED MODEL

ResNet-50
PRETRAINED MODEL

ONNX Converter
MODEL CONVERTER

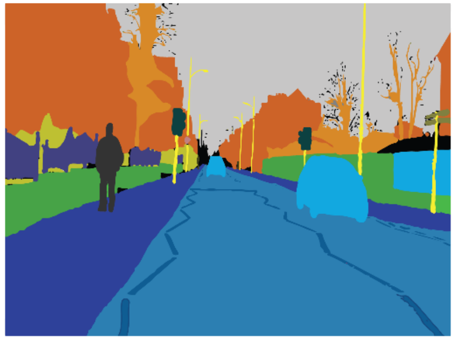
Caffe
IMPORTER

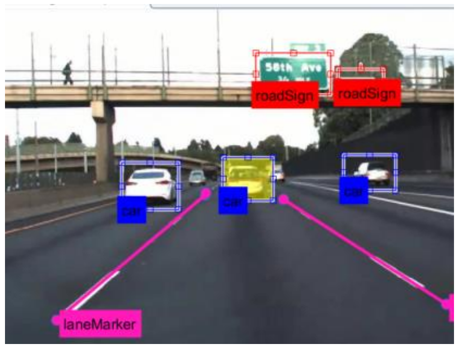
GoogLeNet
PRETRAINED MODEL

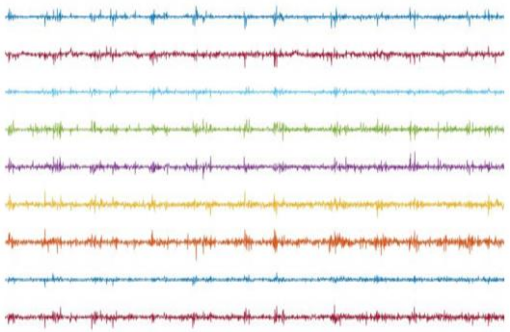
TensorFlow-Keras
IMPORTER

Inception-v3
MODELS

- Labeled public datasets



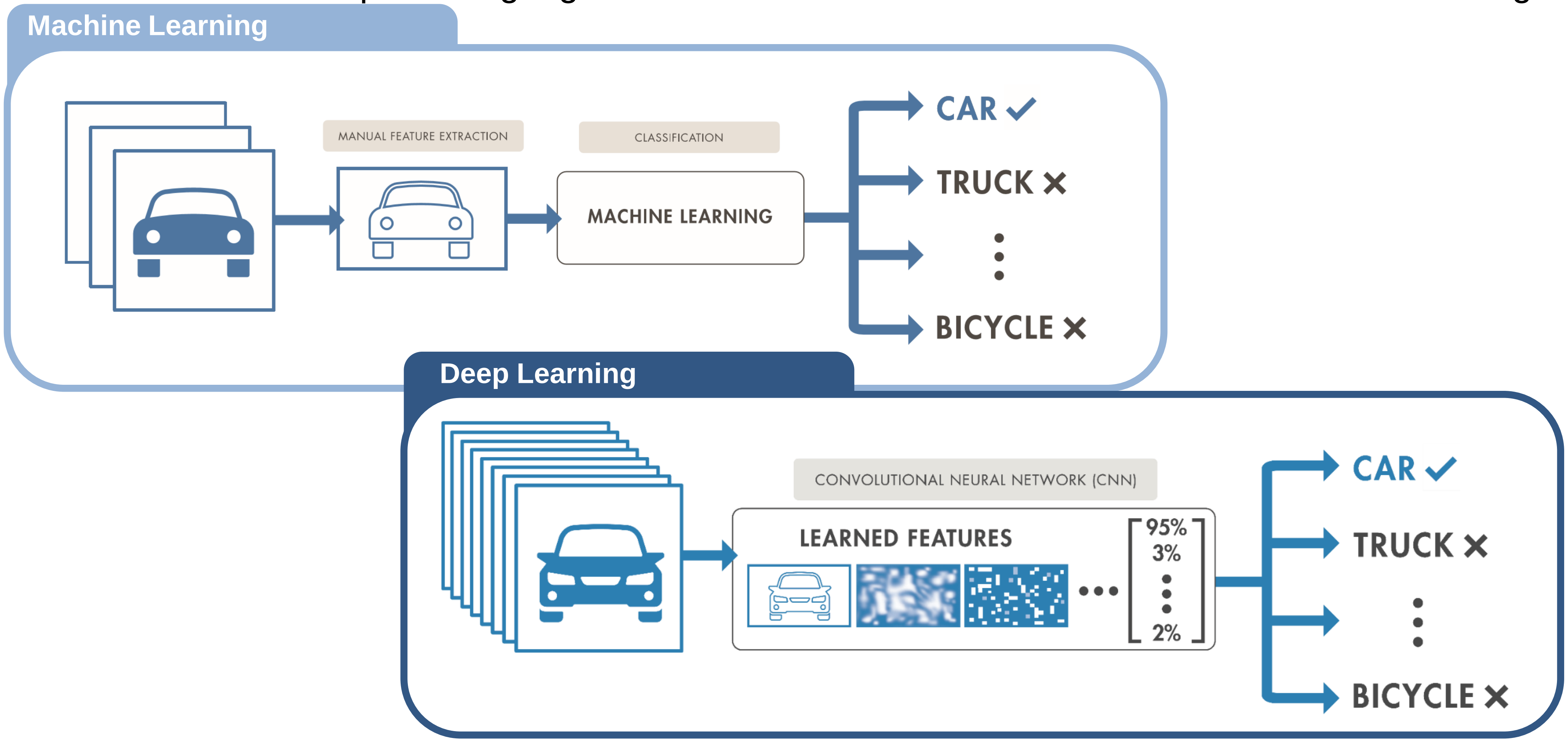




Machine Learning vs Deep Learning

Deep learning performs **end-to-end learning** by learning **features, representations and tasks** directly from **images, text and sound**

Deep learning algorithms also **scale with data –** traditional machine learning **saturates**



Deep Learning Workflow

Access and Explore Data	Label and Preprocess Data	Develop Predictive Models	Integrate Models with Systems

Access and Explore Data

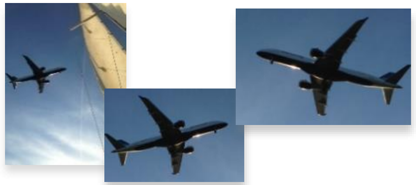
Files

Databases

Sensors

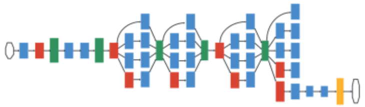
Label and Preprocess Data

Data Augmentation/ Transformation



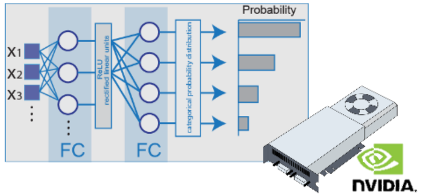
Labeling Automation

Import Reference Models



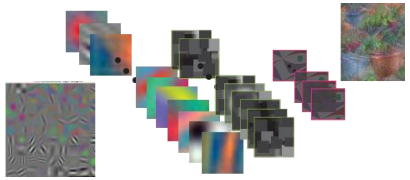
Develop Predictive Models

Hardware-Accelerated Training



Hyperparameter Tuning

Network Visualization



Integrate Models with Systems

Desktop Apps

Enterprise Scale Systems

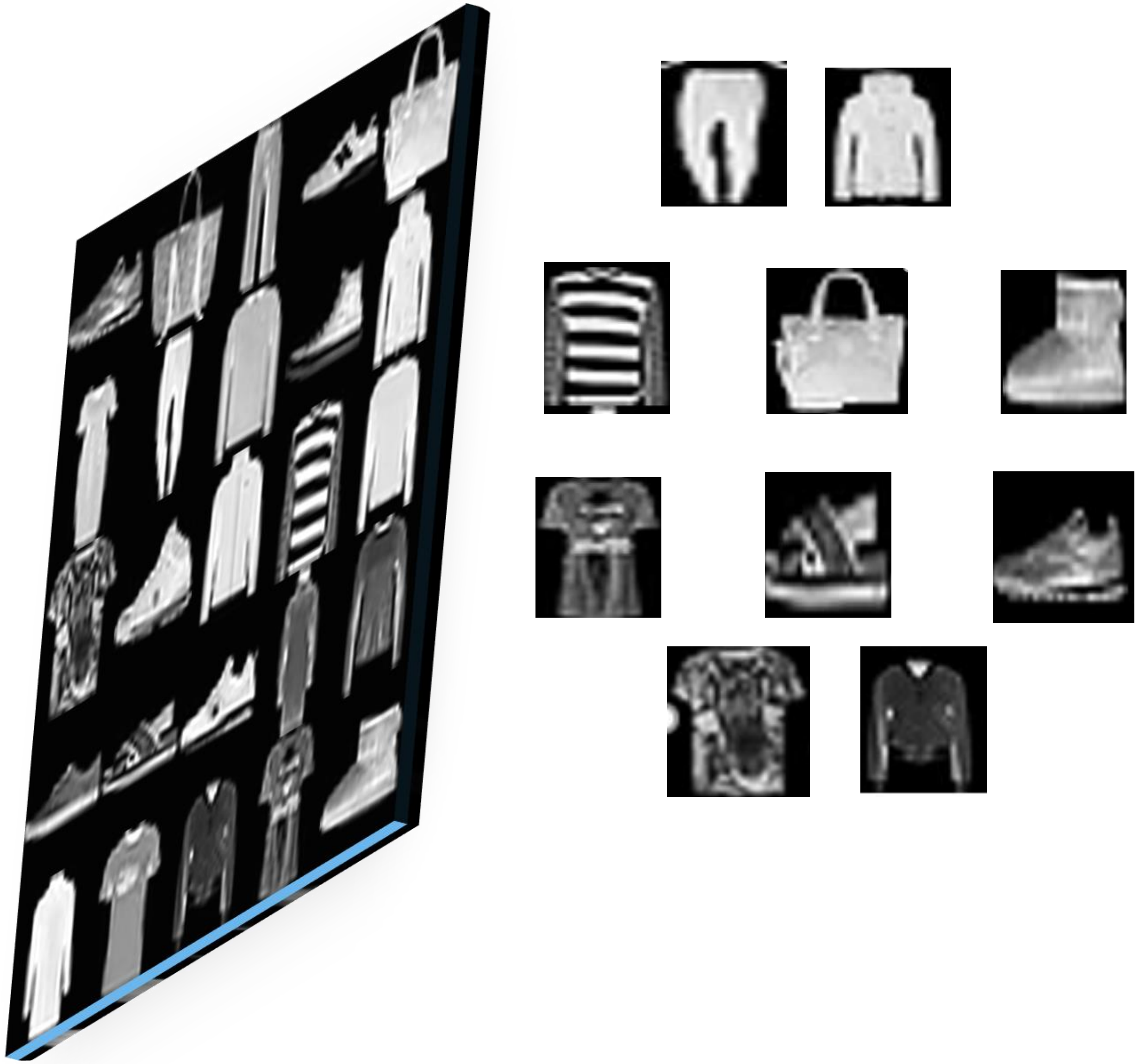
Embedded Devices and Hardware

Agenda

- Why deep learning?
- Fashion MNIST: The “Hello, World!” of deep learning
- Transfer learning with CNNs
- (optional) Semantic segmentation
- (optional) Deep learning with time series data
- Ground Truth Labeling for datasets
- Everything else in deep learning…

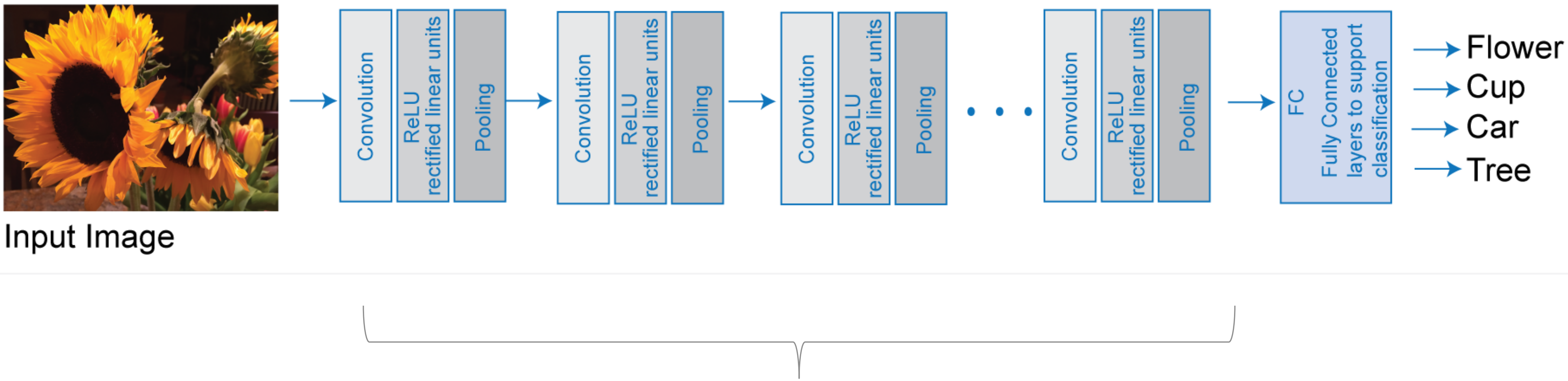
Fashion-MNIST Dataset

What?	A collection if items such as bags, shoes, etc.
Why?	Benchmark machine learning algorithms
How many?	60,000 training images 10,000 test images
Best results?	96.3% accuracy



Sources: https://github.com/zalandoresearch/fashion-mnist

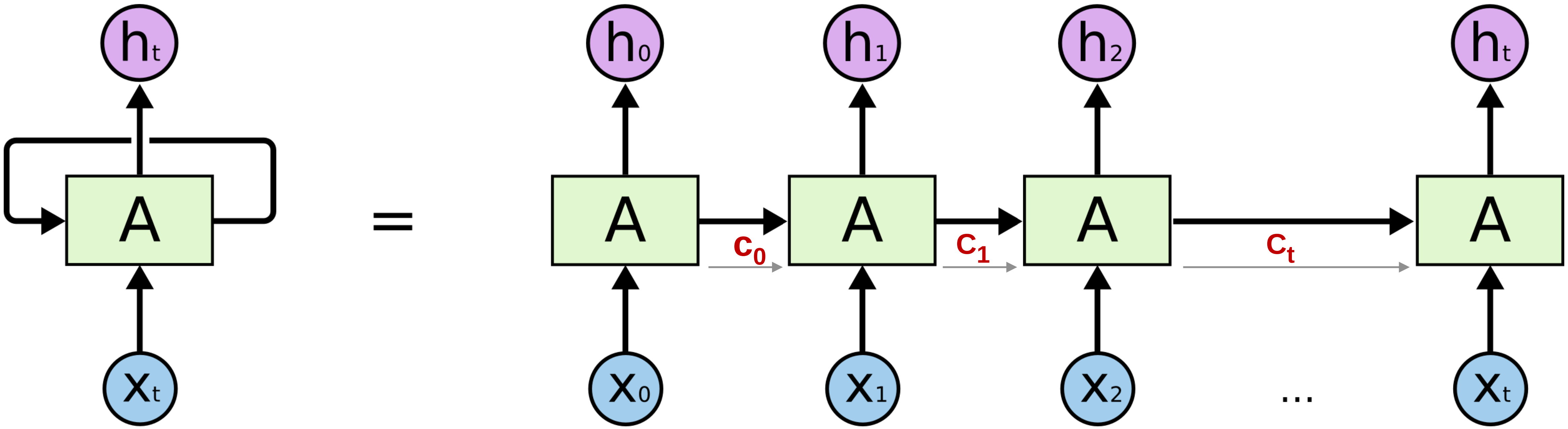
Convolutional Neural Networks (CNN)



"Deep" in deep learning refers to number of layers

Long Short Term Memory Networks

- Recurrent Neural Network that carries a memory cell throughout the process
- Sequence Problems



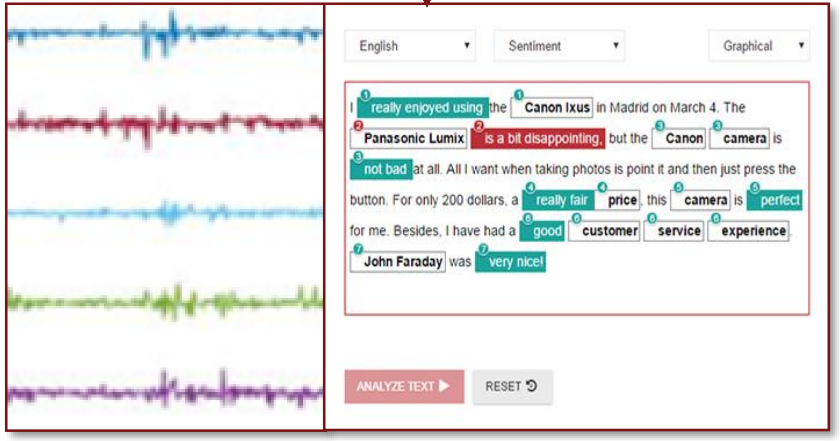
Types of Datasets

Numeric Data

ID	WC_TA	RE_TA	EBIT_TA	MVE_BVTD	S_TA	Industry	Rating
62394	0.013	0.104	0.036	0.447	0.142	3	BB
48608	0.232	0.335	0.062	1.969	0.281	8	A
42444	0.311	0.367	0.074	1.935	0.366	1	A
48631	0.194	0.263	0.062	1.017	0.228	4	BBB
43768	0.121	0.413	0.057	3.647	0.466	12	AAA
39255	-0.117	-0.799	0.01	0.179	0.082	4	CCC
62236	0.087	0.158	0.049	0.816	0.324	2	BBB
39354	0.005	0.181	0.034	2.597	0.388	7	AA
40326	0.47	0.752	0.07	11.596	1.12	8	AAA
51681	0.11	0.337	0.045	3.835	0.812	4	AAA

Machine Learning or LSTM

Time Series/ Text Data



LSTM or CNN

Image Data

CNN

Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

(optional) Semantic segmentation

(optional) Deep learning with time series data

Ground Truth Labeling for datasets

Everything else in deep learning…

Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

(optional) Semantic segmentation

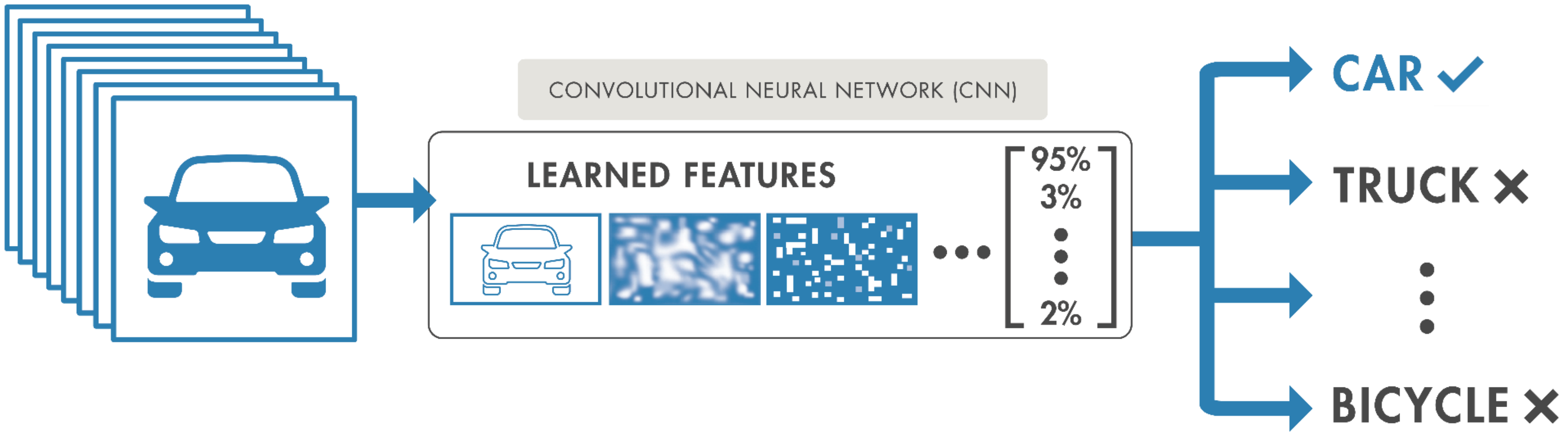
(optional) Deep learning with time series data

Ground Truth Labeling for datasets

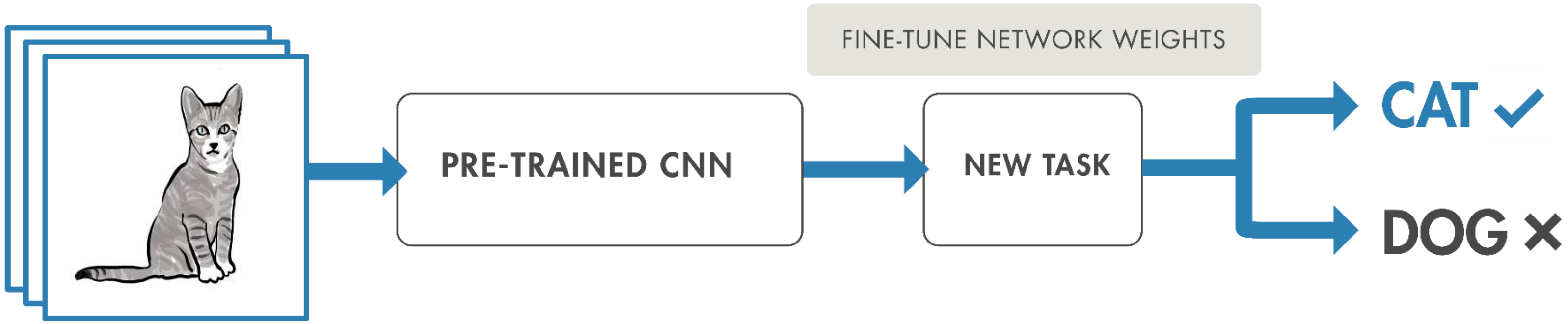
Everything else in deep learning…

Two Approaches for Deep Learning

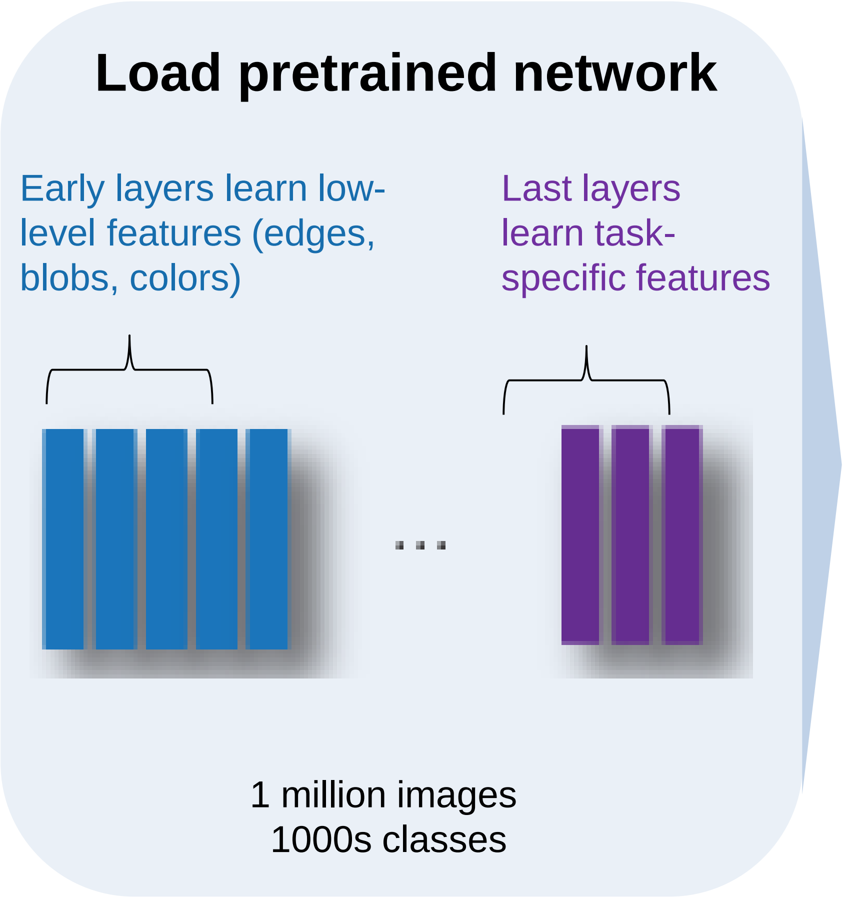
1. Train a Deep Neural Network from Scratch



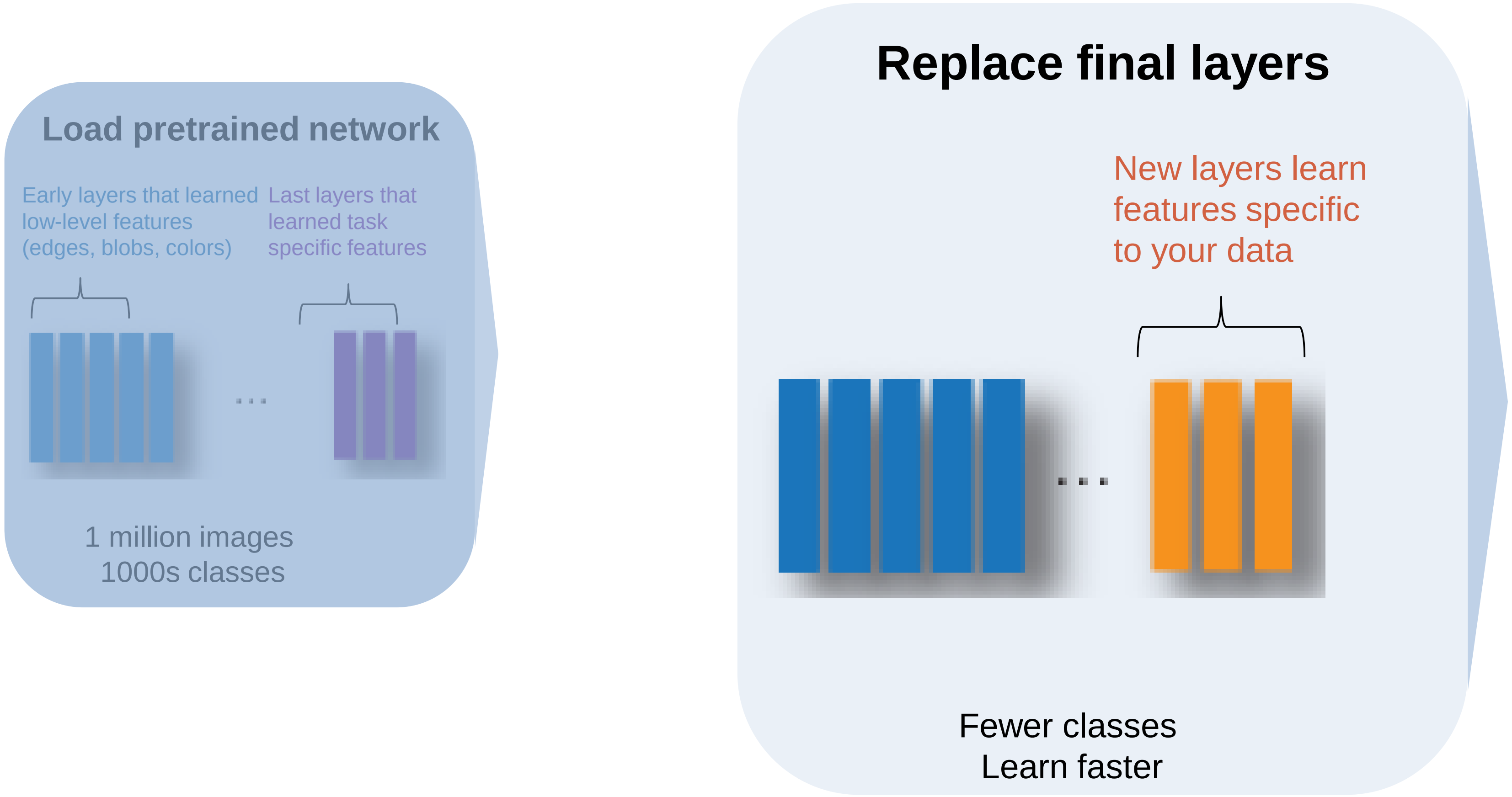
2. Fine-tune a pre-trained model (transfer learning)



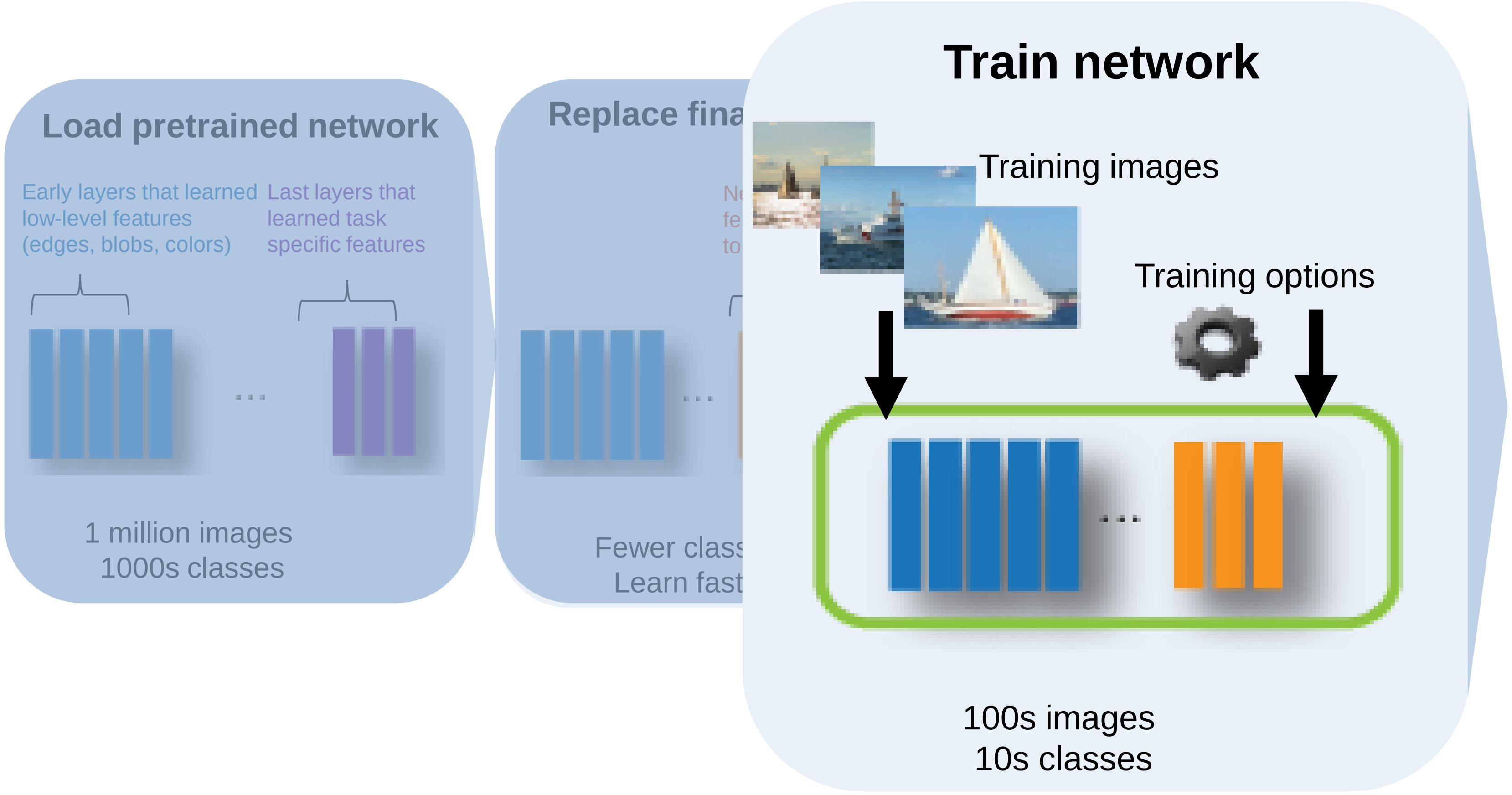
Transfer Learning Workflow



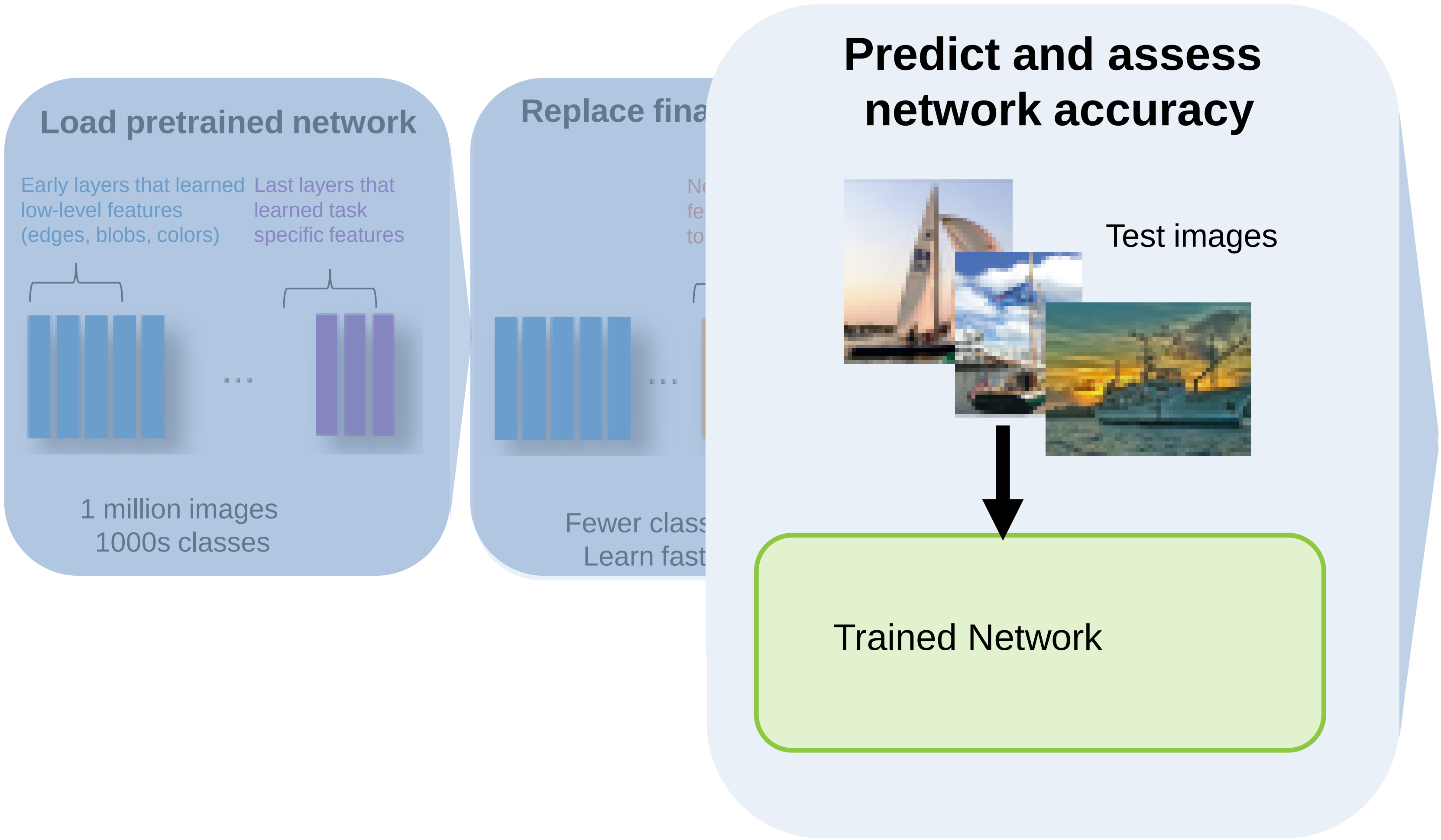
Transfer Learning Workflow



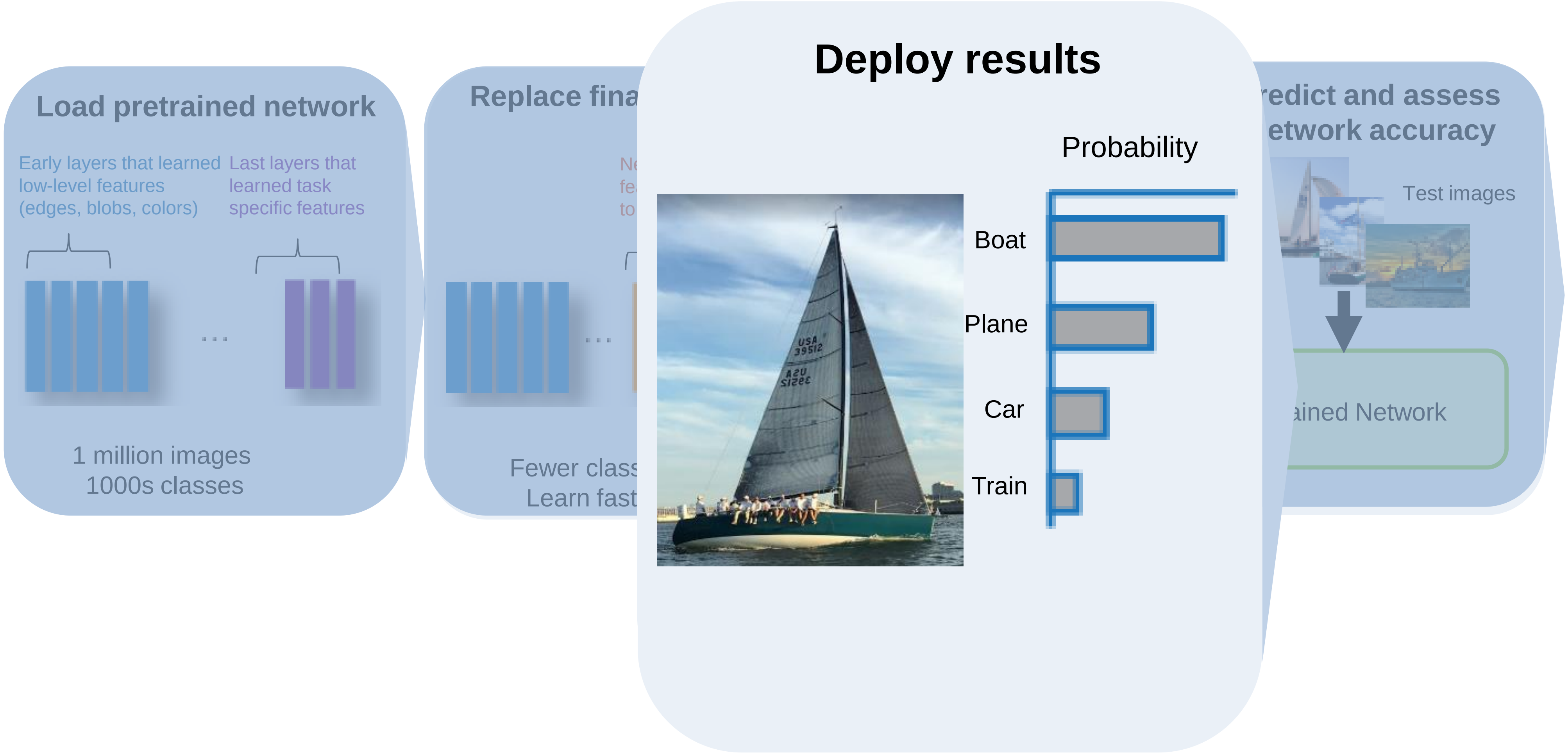
Transfer Learning Workflow



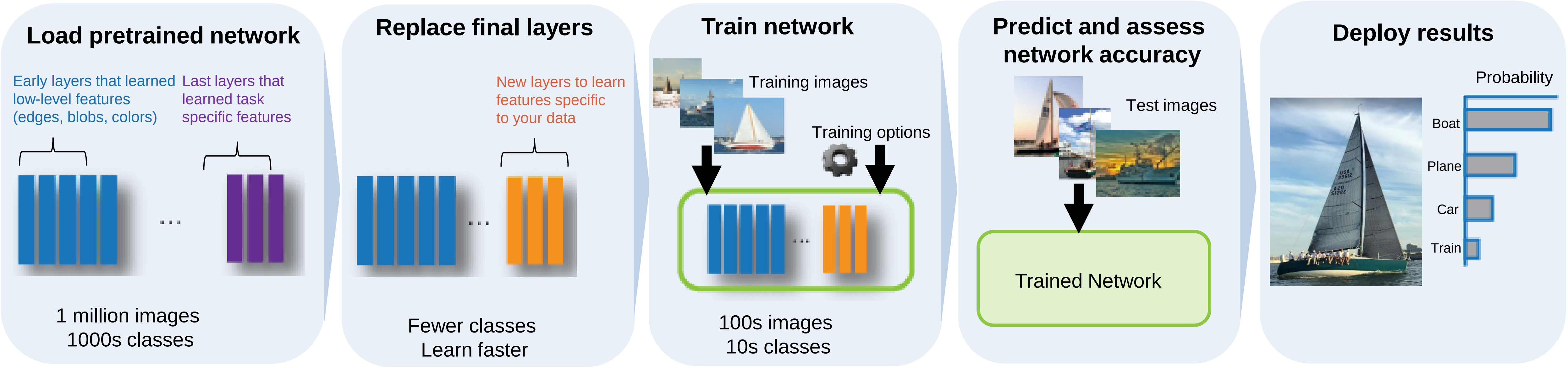
Transfer Learning Workflow



Transfer Learning Workflow



Transfer Learning Workflow



Example: Food classifier using deep transfer learning

Caesar salad →

French fries →

Burgers →

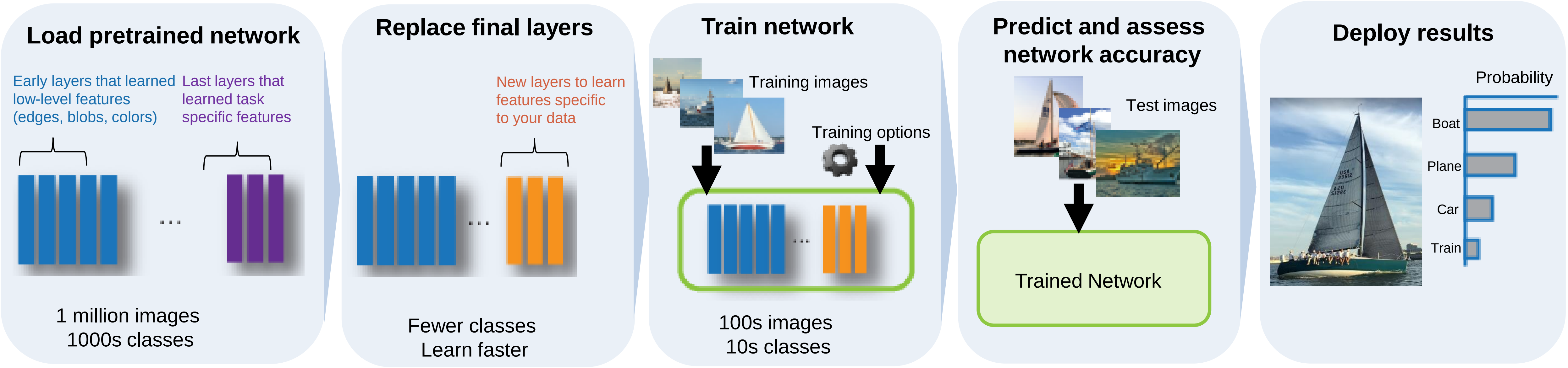
Pizza →

Sushi →

5 Category Classifier


Deep Learning Demo
Image Classification

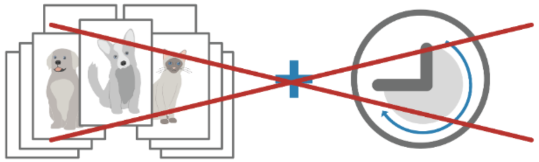
Transfer Learning Workflow



Why Perform Transfer Learning

- Requires less data and training time

- Reference models (like AlexNet, VGG-16, VGG-19, Inception-v3) are great feature extractors



- Leverage best network types from top researchers
  (list of all models)

AlexNet PRETRAINED MODEL	**VGG-16** PRETRAINED MODEL	**ResNet-50** PRETRAINED MODEL	**ONNX Converter** MODEL CONVERTER
Caffe IMPORTER	**GoogLeNet** PRETRAINED MODEL	**TensorFlow-Keras** IMPORTER	**Inception-v3** MODELS

Import the Latest Models for Transfer Learning

Pretrained Models*

- AlexNet
- VGG-16
- VGG-19
- GoogLeNet
- Inception-v3
- ResNet-18
- ResNet-50
- ResNet-101
- Inception-resnet-v2
- SqueezeNet
- DenseNet-201

Import Models from Frameworks

- Caffe Model Importer
- TensorFlow-Keras Model Importer
- ONNX Converter (Import and Export)

AlexNet
PRETRAINED MODEL

VGG-16
PRETRAINED MODEL

ResNet-50
PRETRAINED MODEL

ResNet-101
PRETRAINED MODEL

Caffe
IMPORTER

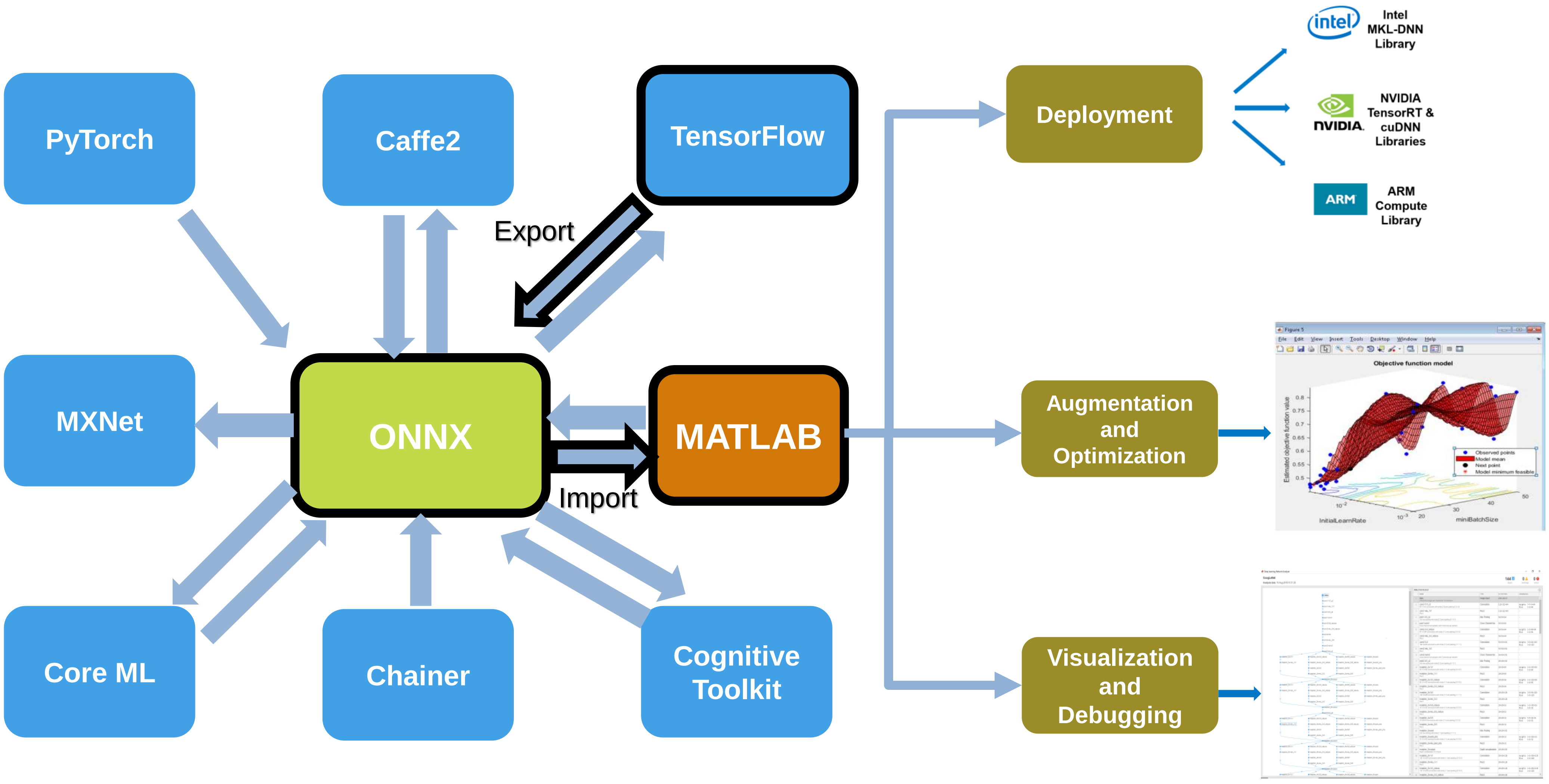
GoogLeNet
PRETRAINED MODEL

TensorFlow-Keras
IMPORTER

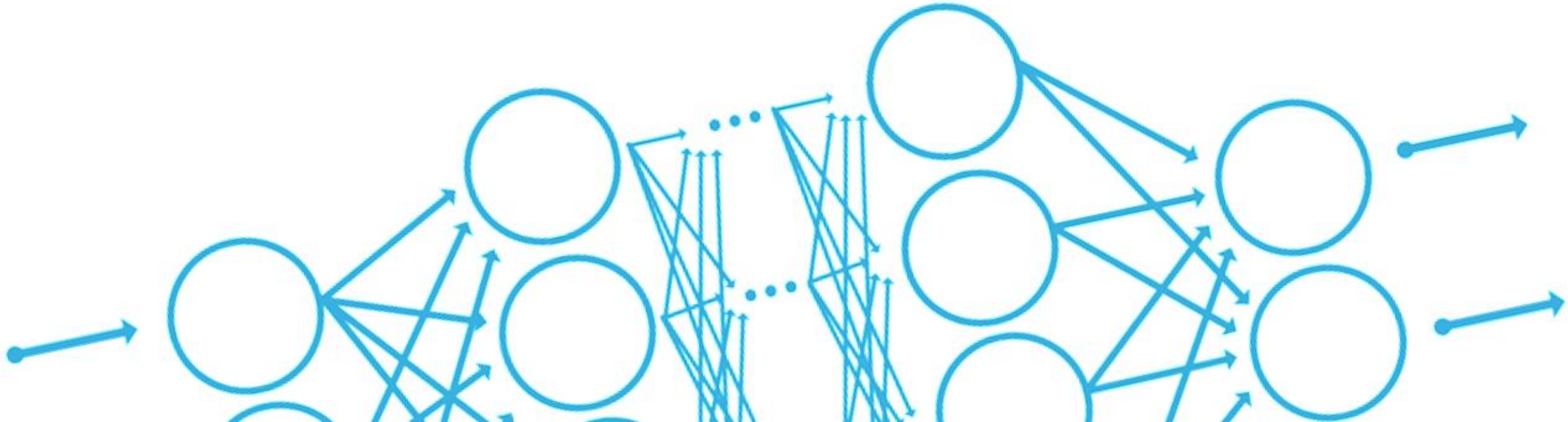
Inception-v3
MODELS

* single line of code to access model

Integration with Other Frameworks

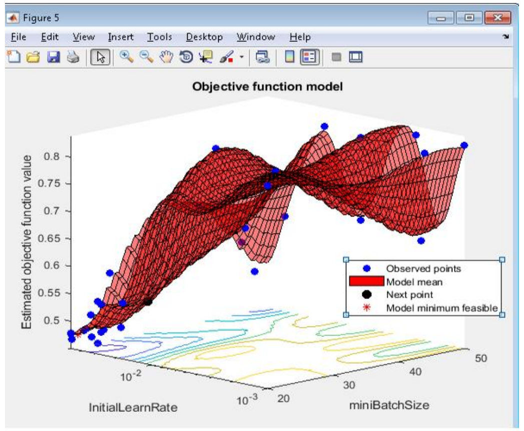


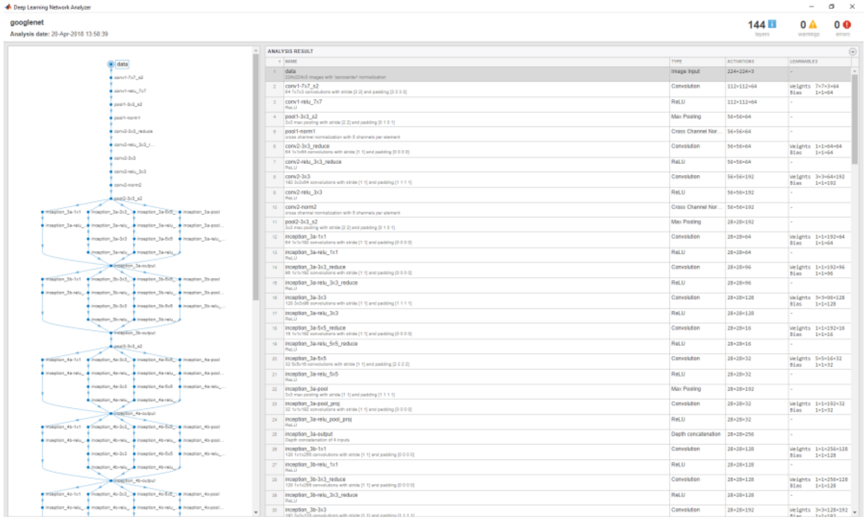
A short recap...

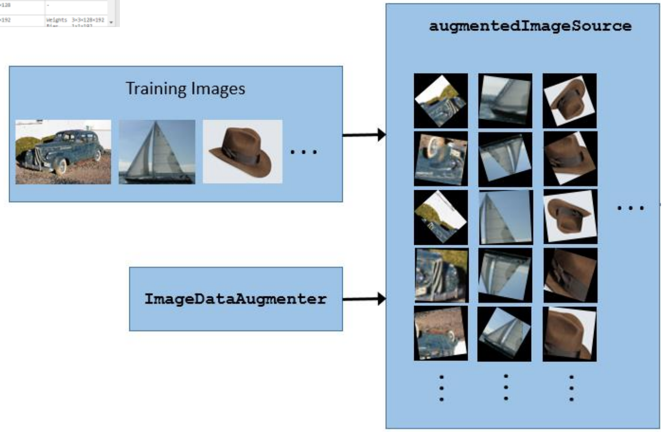


What we covered in this section…

- **Transfer learning**
  - Easily modify existing networks with one-line commands
- **Access large datasets**
  - Using the imageDatastore
- **Visualize and Analyze Networks**
  - Using the network Analyzer
- **Make training faster**
  - Freezing the layers
- **Improve training results**
  - Data augmentation with augmentedImageDataStore
  - Automatic parameter selection with **Bayesian hyperparameter tuning**

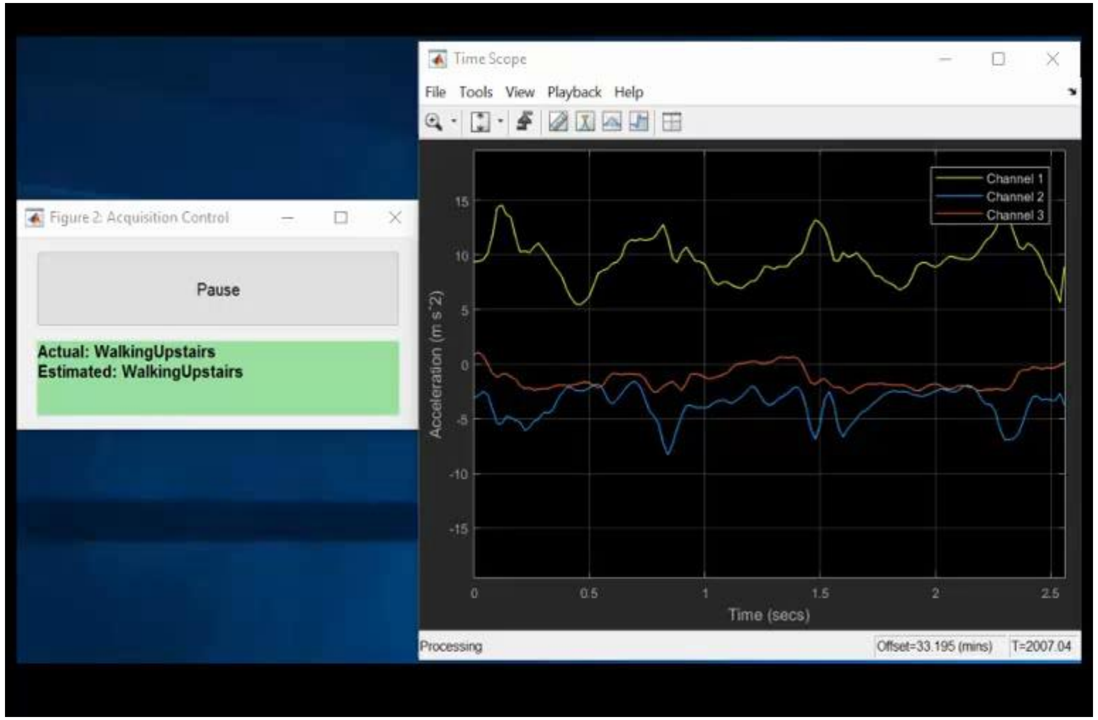






Optional Demo: Image or Signal Data

Semantic Segmentation using SegNet



Signal Classification using LSTMs

Agenda

- Why deep learning?
- Fashion MNIST: The “Hello, World!” of deep learning
- Transfer learning with CNNs
- (optional) Semantic segmentation
- (optional) Deep learning with time series data
- Ground Truth Labeling for datasets
- Everything else in deep learning…

Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

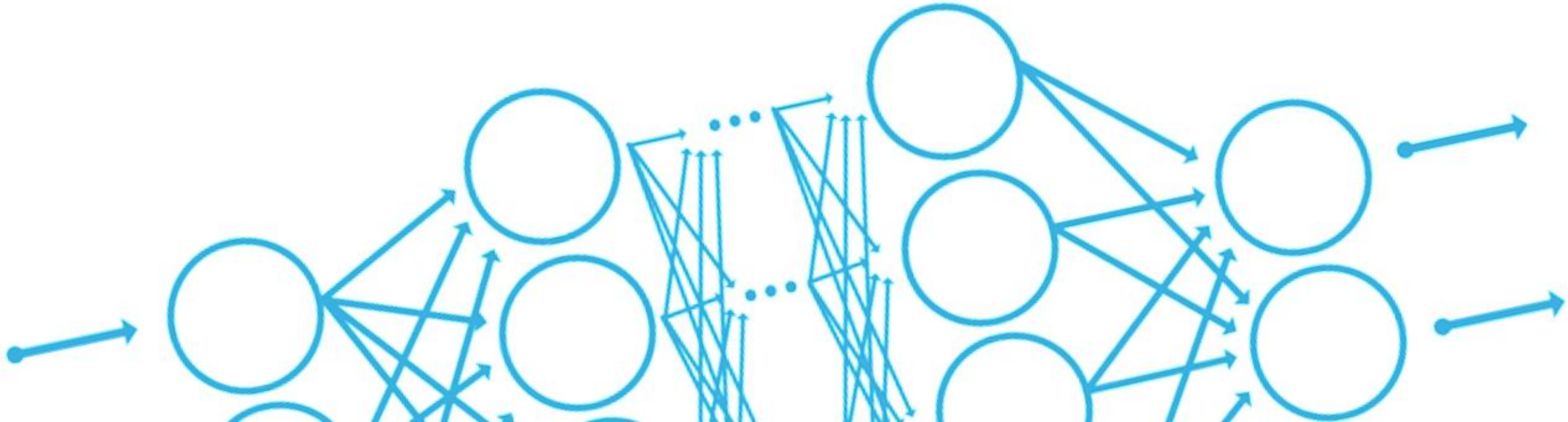
(optional) Semantic segmentation

(optional) Deep learning with time series data

Ground Truth Labeling for datasets

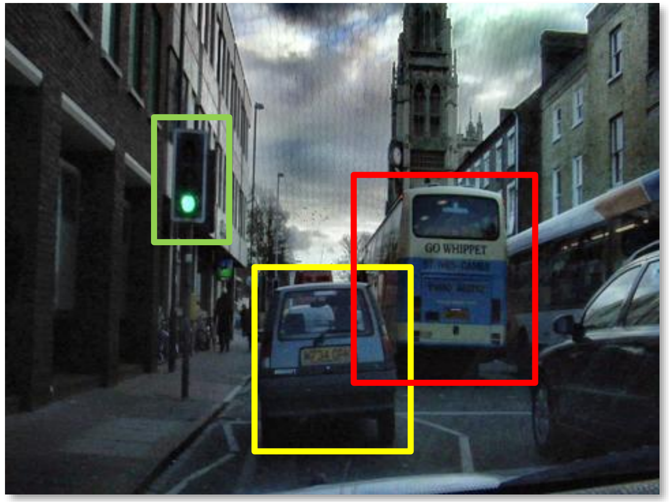
Everything else in deep learning…

What is semantic segmentation?



Original Image

ROI detection



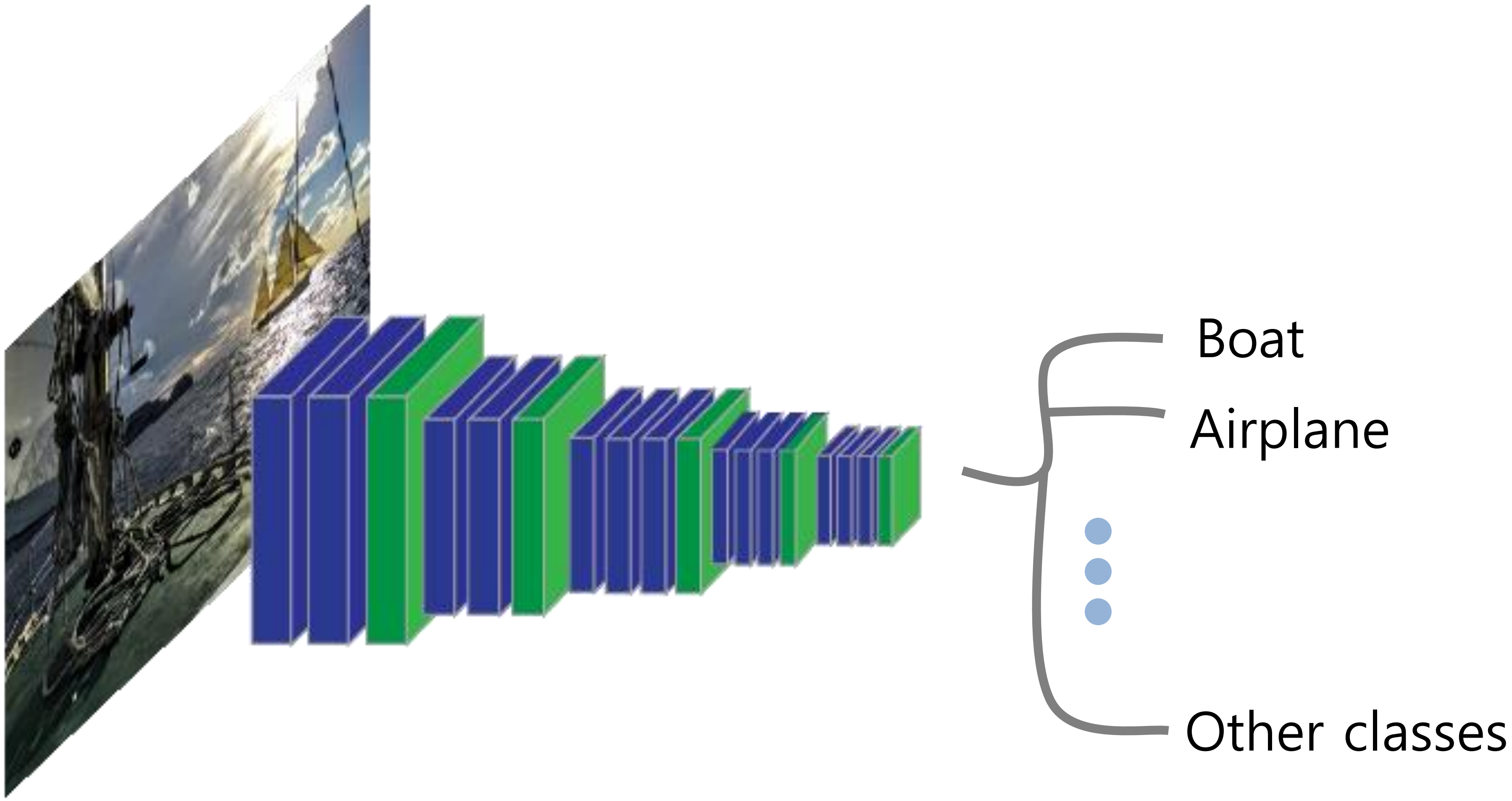
Pixel classification

Semantic Segmentation

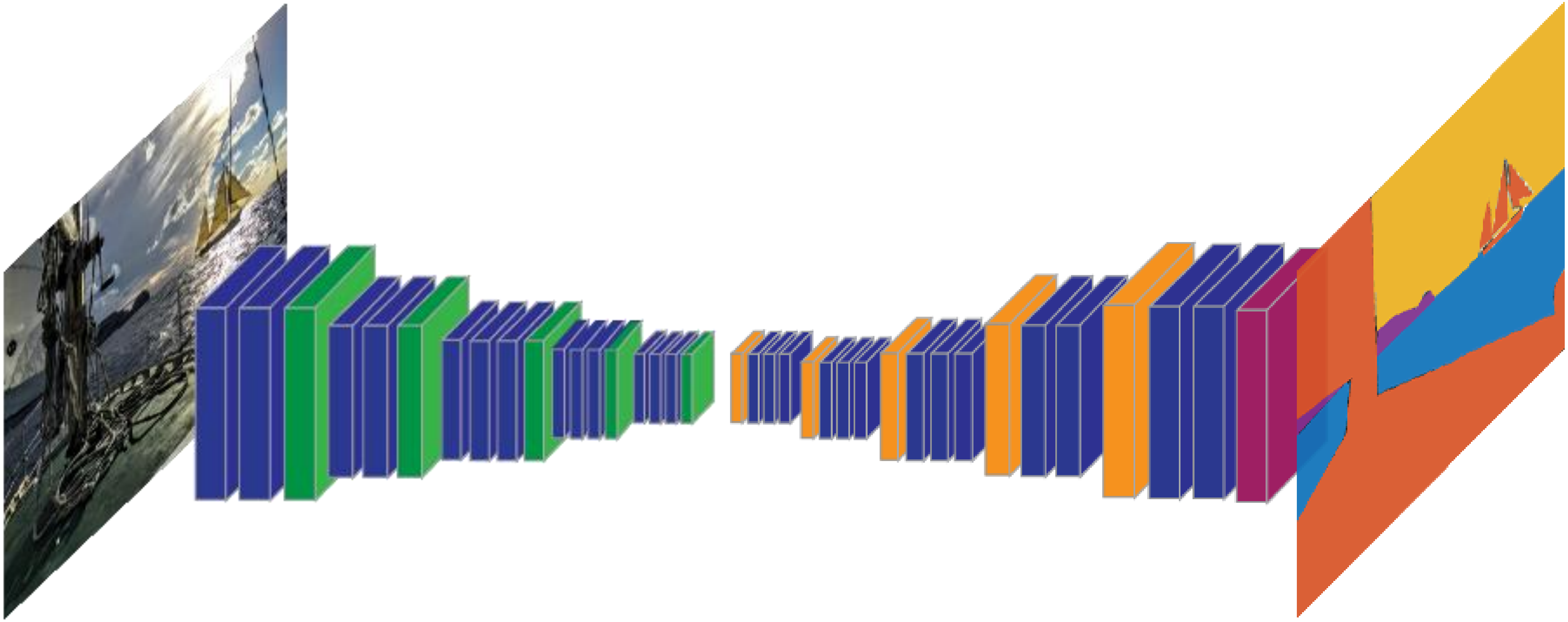
CamVid Dataset

1. Segmentation and Recognition Using Structure from Motion Point Clouds, ECCV 2008
2. Semantic Object Classes in Video: A High-Definition Ground Truth Database ,Pattern Recognition Letters

Semantic Segmentation Network

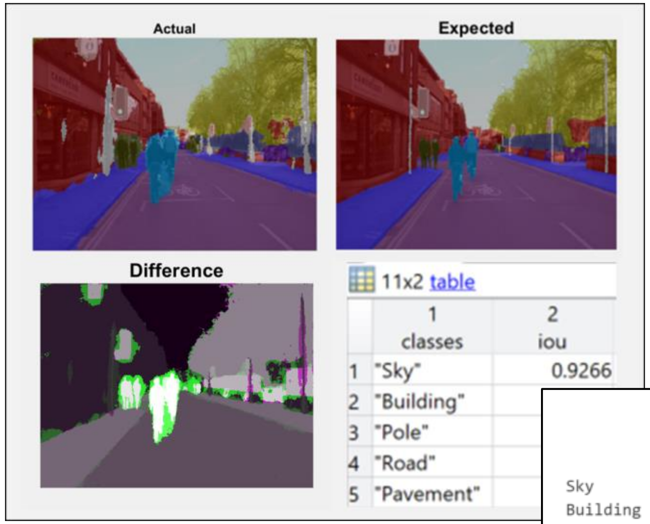


Semantic Segmentation Network

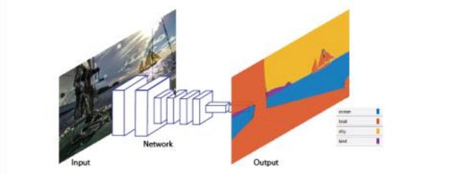


Useful Tools for Semantic Segmentation

- **Automatically create network structures**
  - Using **`segnetLayers`** and **`fcnLayers`**
- **Handle pixel labels**
  - Using the **`pixelLabelImageDatastore`** and **`pixelLabelDatastore`**
- **Evaluate network performance**
  - Using `evaluateSemanticSegmentation`
- **Examples and tutorials to learn concepts**



	Accuracy	IoU	MeanBFScore
Sky	0.93544	0.89279	0.88239
Building	0.79978	0.75543	0.59861
Pole	0.73166	0.18361	0.51426
Road	0.93644	0.90663	0.7086
Pavement	0.90624	0.72932	0.70585
Tree	0.86587	0.73694	0.67097
SignSymbol	0.76118	0.35339	0.44175
Fence	0.83258	0.49648	0.50265
Car	0.90961	0.75263	0.64837
Pedestrian	0.83751	0.35409	0.46796
Bicyclist	0.84156	0.5472	0.46933


Semantic Segmentation Basics R2018a

Segmentation is essential for image analysis tasks. Semantic segmentation describes the process of associating each pixel of an image with a class label, (such as flower, person, road, sky, ocean, or car).

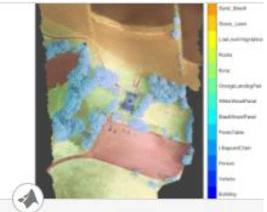
Applications for semantic segmentation include:

- Autonomous driving
- Industrial inspection
- Classification of terrain visible in satellite imagery
- Medical imaging analysis

Train a Semantic Segmentation Network

The steps for training a semantic segmentation network are as follows:

1. Analyze Training Data for Semantic Segmentation
2. Create a Semantic Segmentation Network
3. Train A Semantic Segmentation Network
4. Evaluate and inspect the results of semantic segmentation



Semantic Segmentation of Multispectral Images Using Deep Learning

Train a U-Net convolutional neural network to perform semantic segmentation of a multispectral image with seven channels: three

Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

(optional) Semantic segmentation

(optional) Deep learning with time series data

Ground Truth Labeling for datasets

Everything else in deep learning…

Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

(optional) Semantic segmentation

(optional) Deep learning with time series data

Ground Truth Labeling for datasets

Everything else in deep learning…

I was born in France. I speak ___________ ?

I was born in France…

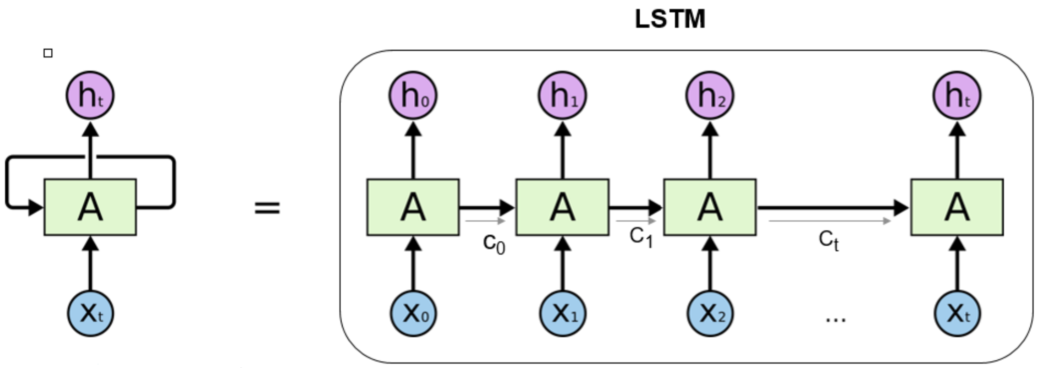
[2000 words]

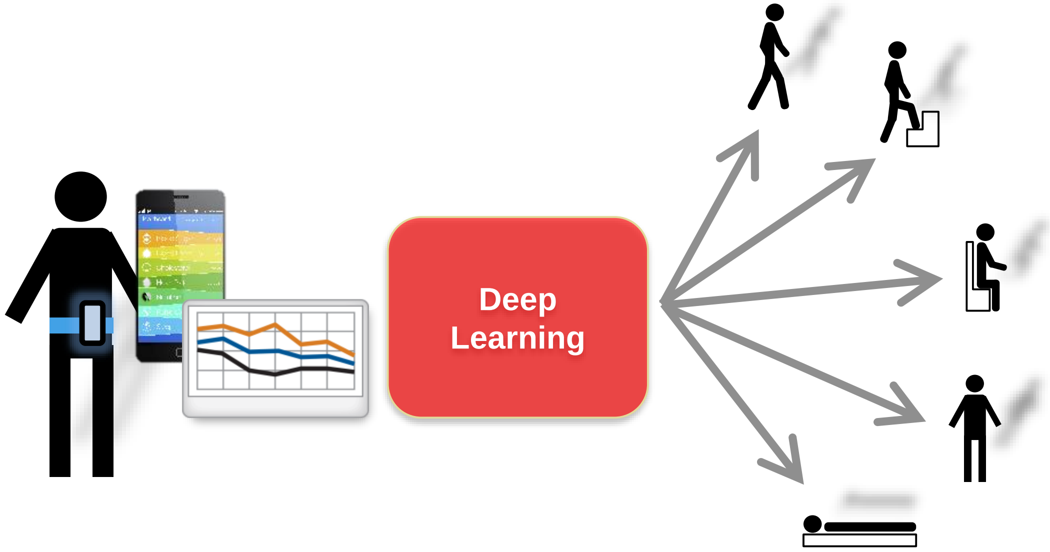
… I speak ___________ ?

Time Series Classification (Human Activity Recognition)

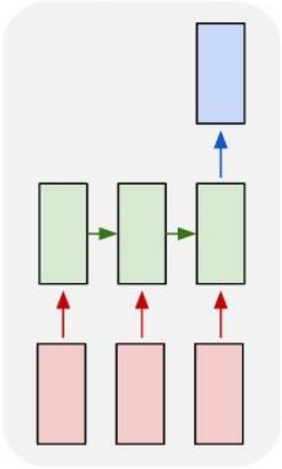
Long short-term memory networks

- Dataset is accelerometer and gyroscope signals captured with a smartphone
- Data is a collection of time series with 9 channels

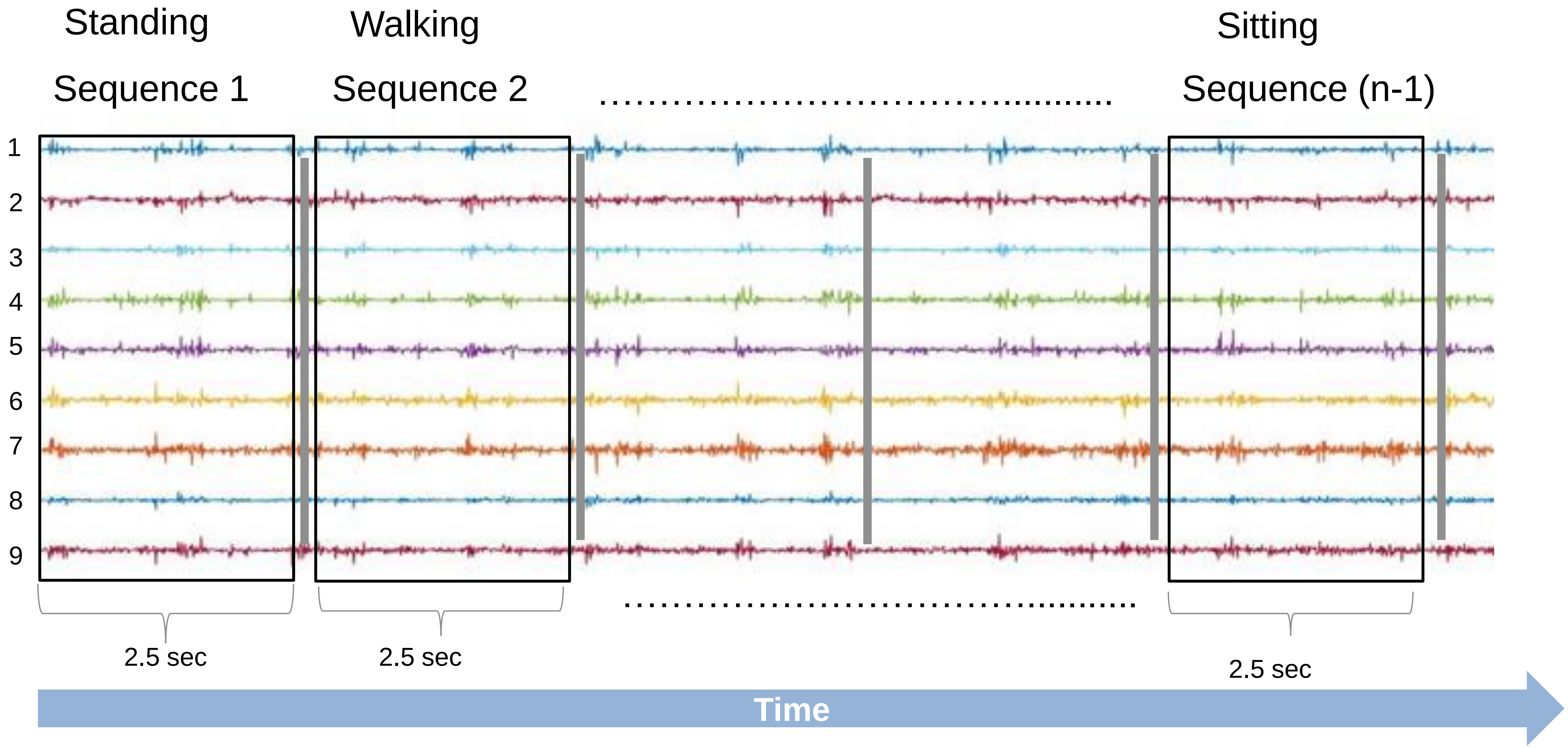




sequence to one



LSTM sequences



Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

(optional) Semantic segmentation

(optional) Deep learning with time series data

Ground Truth Labeling for datasets

Everything else in deep learning…

Agenda

- Why deep learning?
- Fashion MNIST: The "Hello, World!" of deep learning
- Transfer learning with CNNs
- (optional) Semantic segmentation
- (optional) Deep learning with time series data
- Ground Truth Labeling for datasets
- Everything else in deep learning…

“I love to label and preprocess my data”

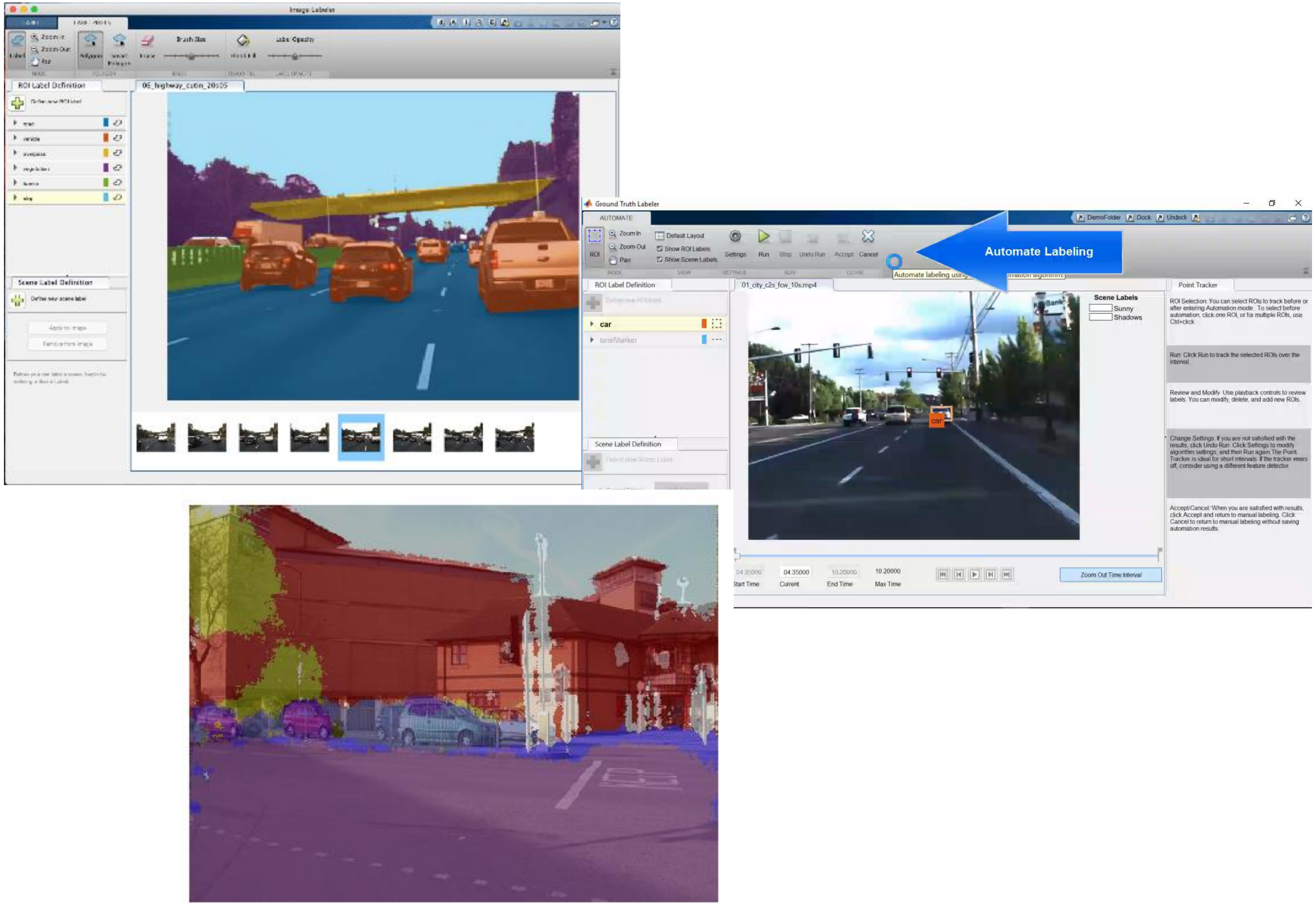
~ Said no engineer, ever.

Ground truth Labeling

“How do I *label* my data?”

New App for Ground Truth Labeling

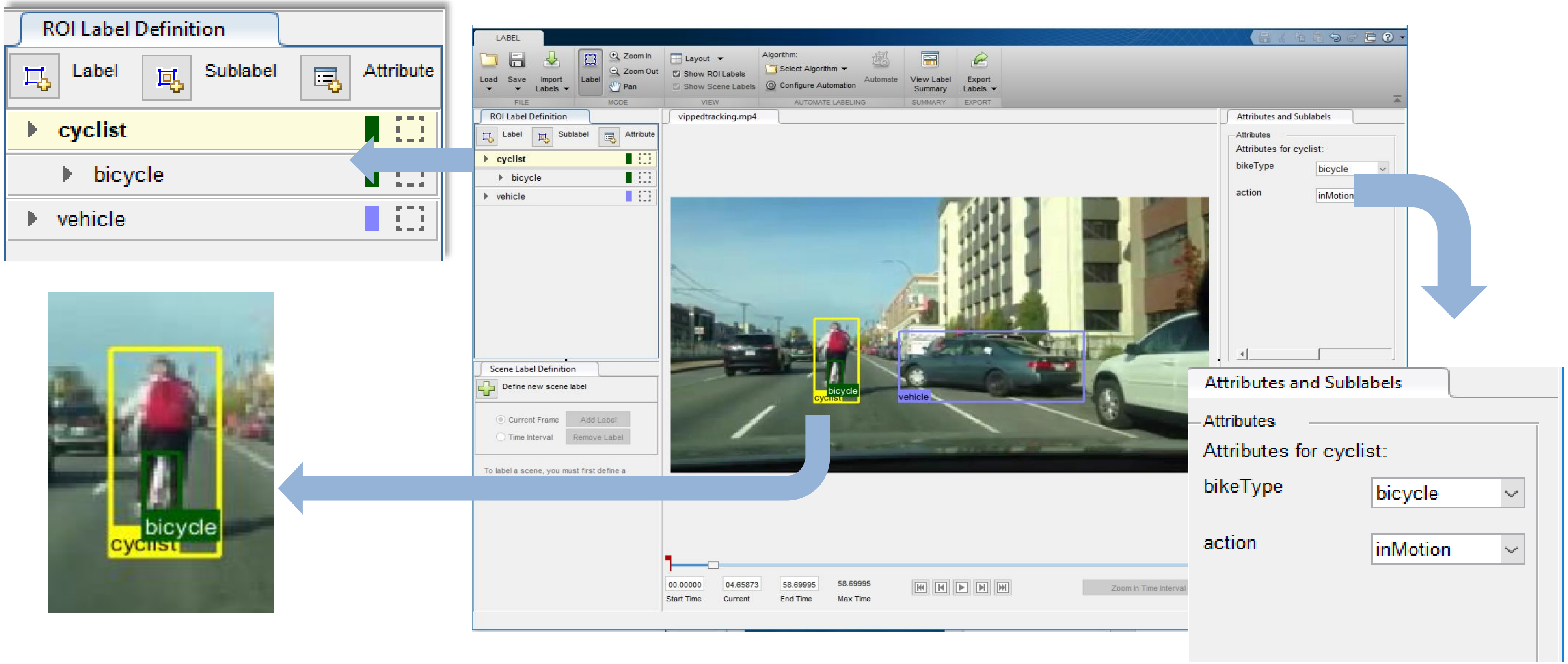
Label pixels and regions for semantic segmentation



Data

Attributes and Sublabels

NEW in R2018a



Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

(optional) Semantic segmentation

(optional) Deep learning with time series data

Ground Truth Labeling for datasets

Everything else in deep learning…

Agenda

Why deep learning?

Fashion MNIST: The “Hello, World!” of deep learning

Transfer learning with CNNs

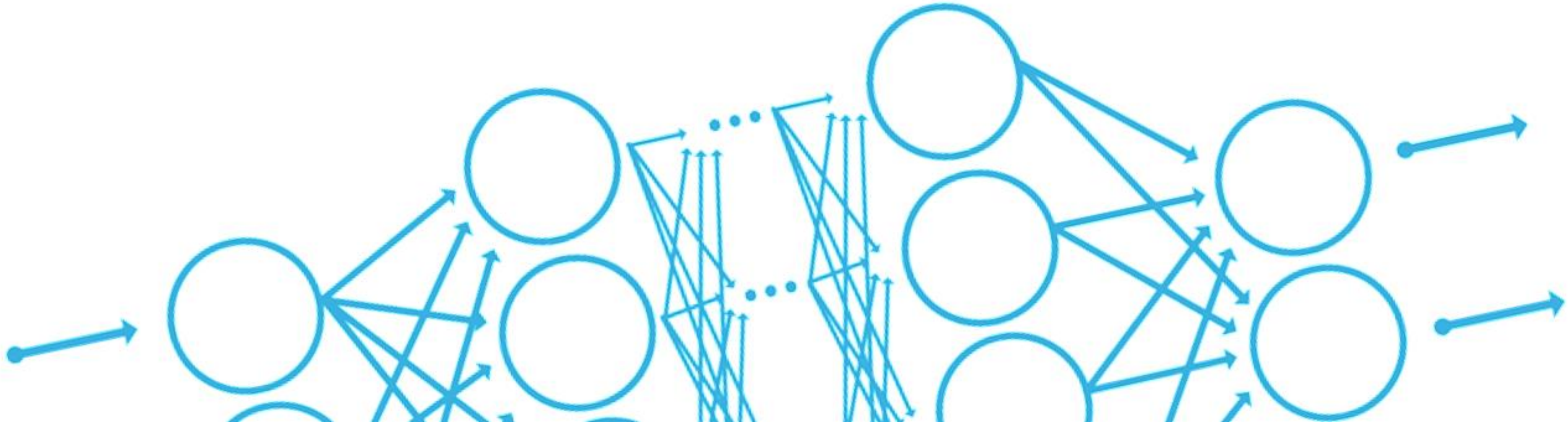
(optional) Semantic segmentation

(optional) Deep learning with time series data

Ground Truth Labeling for datasets

Everything else in deep learning...

Training Performance and Scalability



Deep Learning on CPU, GPU, Multi-GPU and Clusters

How to target?

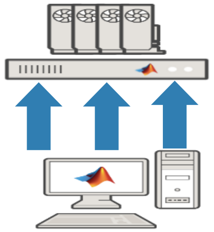
Single
CPU

Single CPU
Single GPU

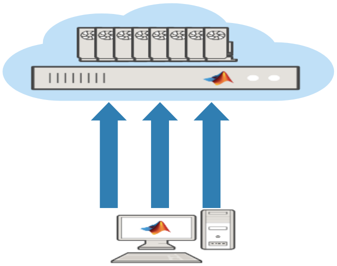
```
opts = trainingOptions('sgdm', ...
      'MaxEpochs', 100, ...
      'MiniBatchSize', 250, ...
      'InitialLearnRate', 0.00005, ...
'ExecutionEnvironment', 'auto' );
```

Single CPU, Multiple GPUs

```
opts = trainingOptions('sgdm', ...
      'MaxEpochs', 100, ...
      'MiniBatchSize', 250, ...
      'InitialLearnRate', 0.00005, ...
'ExecutionEnvironment', 'multi-gpu' );
```



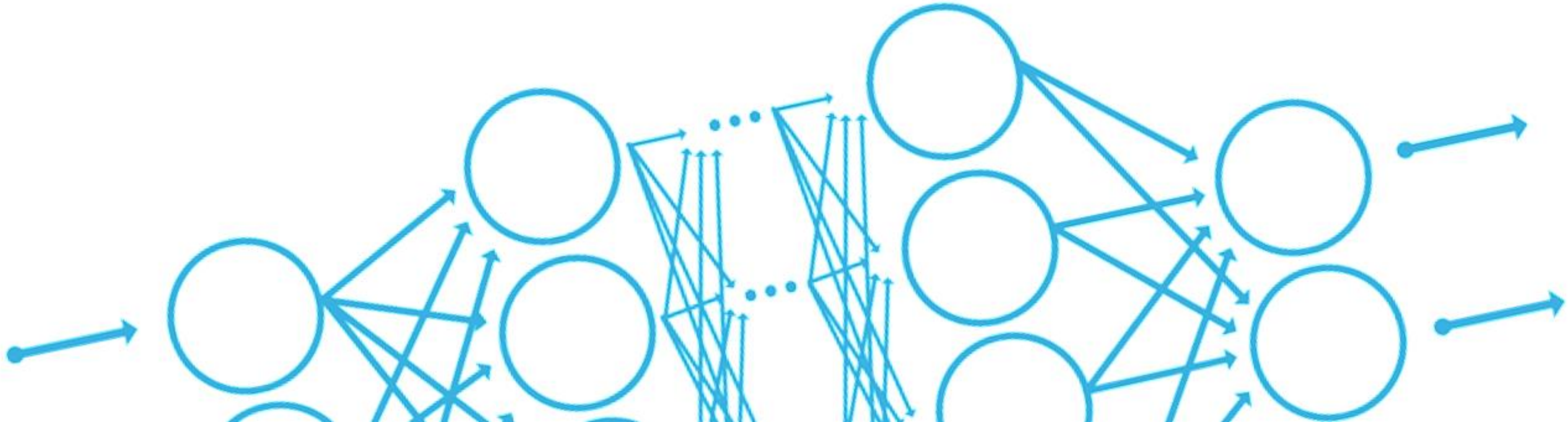
On-prem server with
GPUs



Cloud GPUs
(AWS)

```
opts = trainingOptions('sgdm', ...
      'MaxEpochs', 100, ...
      'MiniBatchSize', 250, ...
      'InitialLearnRate', 0.00005, ...
'ExecutionEnvironment', 'parallel' );
```

Inference Performance and Deployment



Using Coder Products with Deep Learning

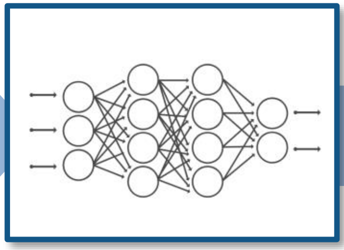
Access Data

Image Acquisition Toolbox

Preprocess

Image Processing Toolbox

Computer Vision System Toolbox

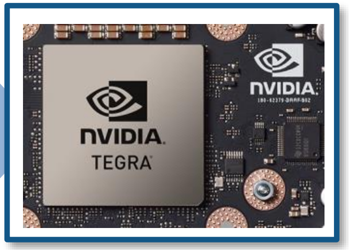


Select Network

Deep Learning Toolbox

Train

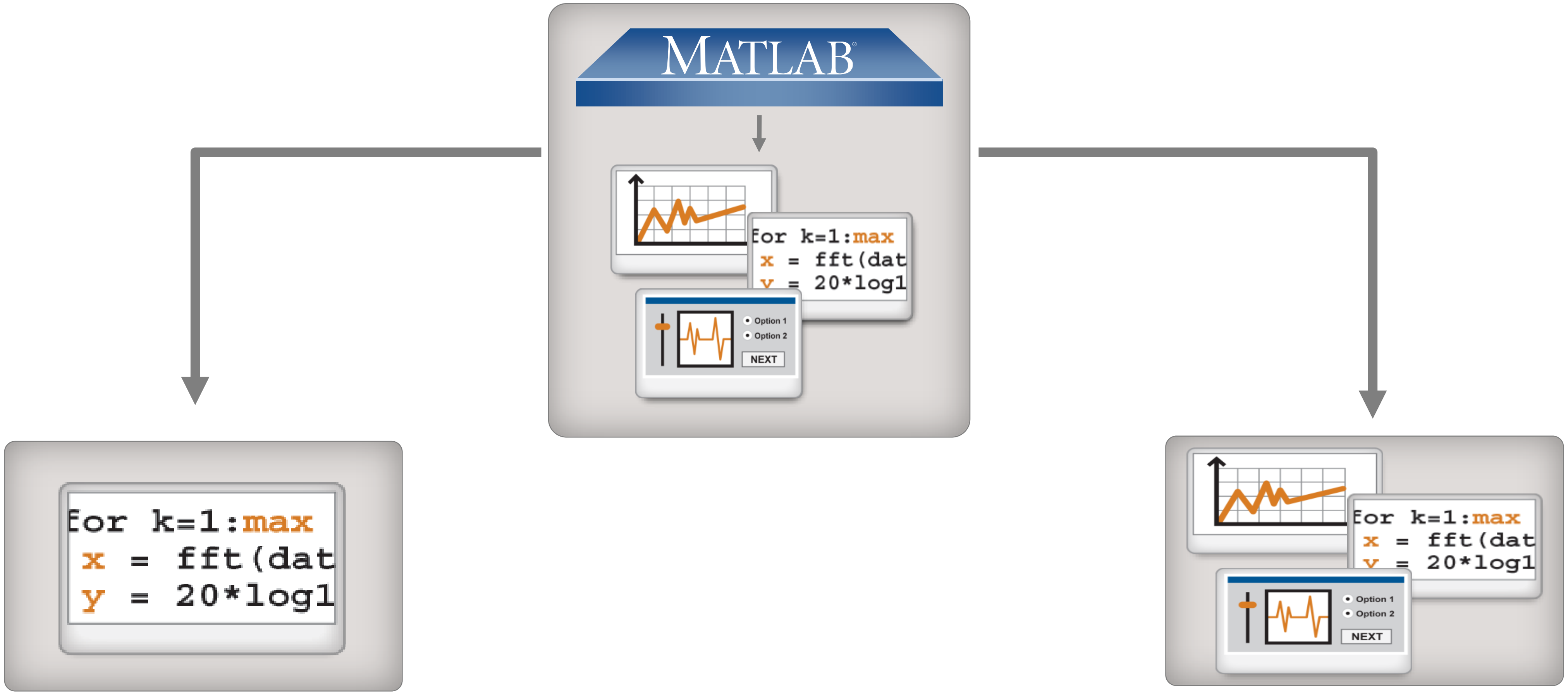
Parallel Computing Toolbox



Deploy

Coder Products

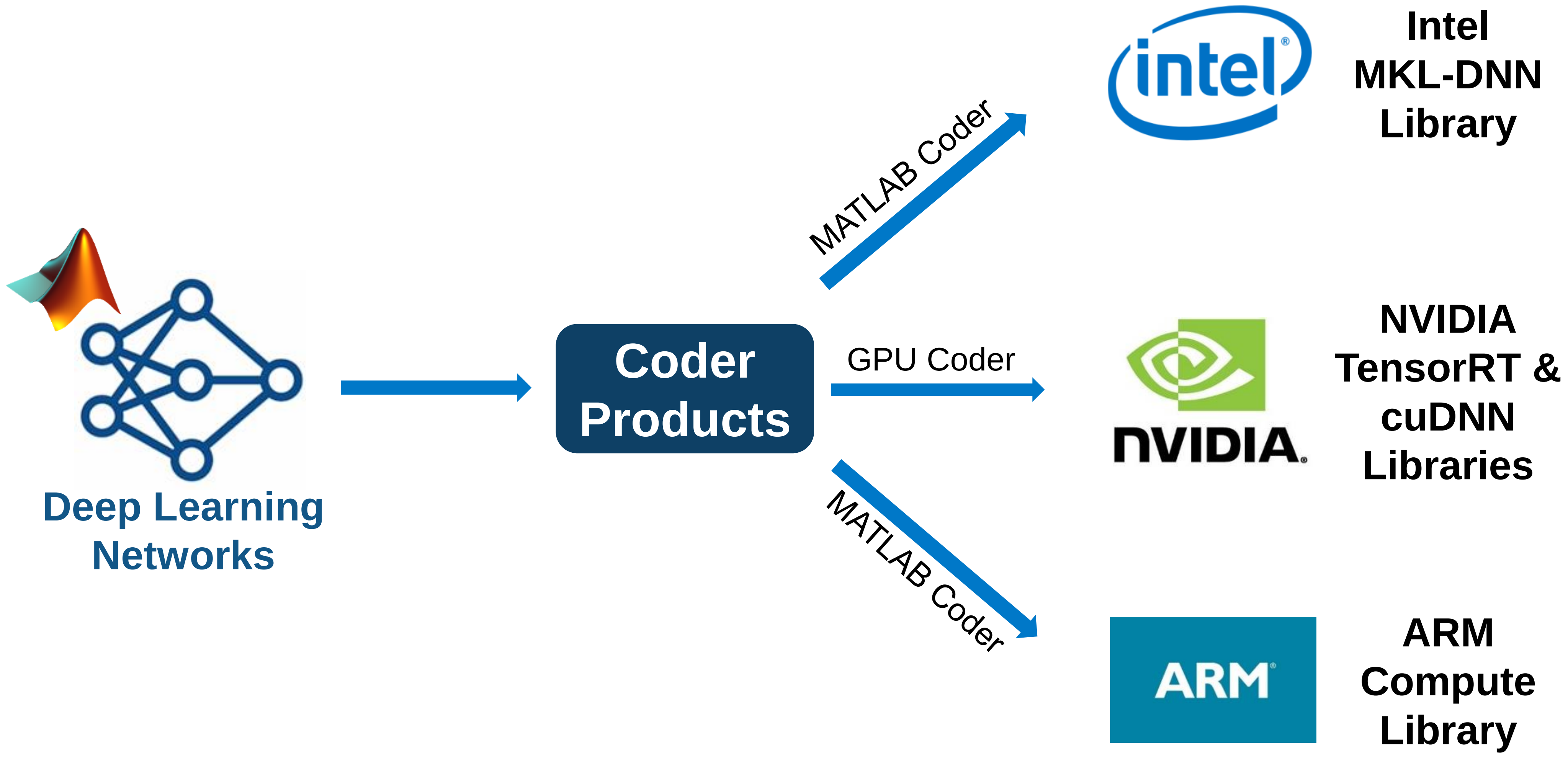
Integrate Deep Learning within Systems



Embedded Hardware

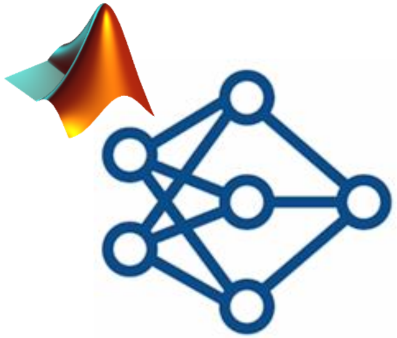
Enterprise Systems

Deploying Deep Learning Models for Inference



Deploying to Various Targets

R2018a



Deep Learning Networks

Coder Products

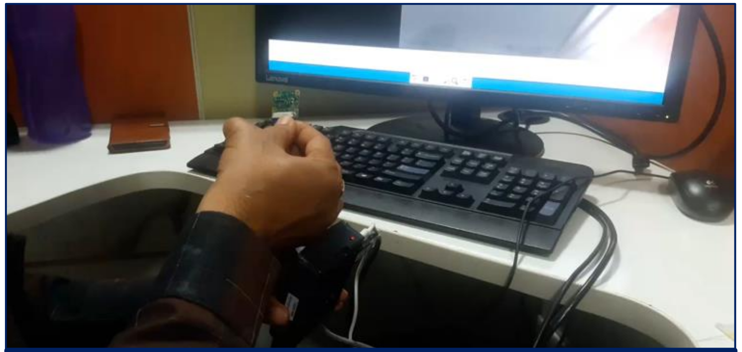
MATLAB Coder

Desktop CPU

GPU Coder

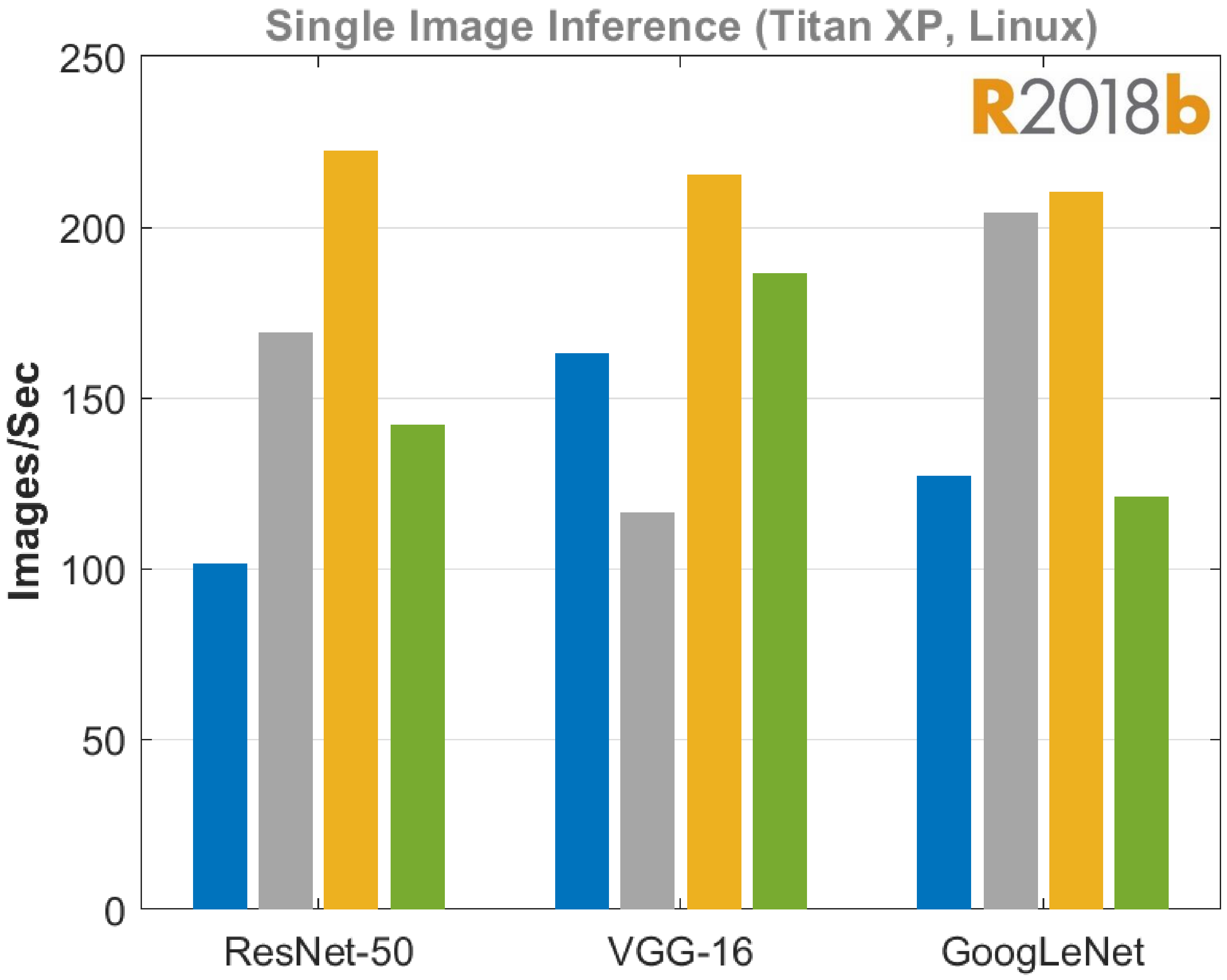
NVIDIA TensorRT & cuDNN Libraries

MATLAB Coder



Raspberry Pi board

With GPU Coder, MATLAB is fast



Faster than TensorFlow, MXNet, and PyTorch

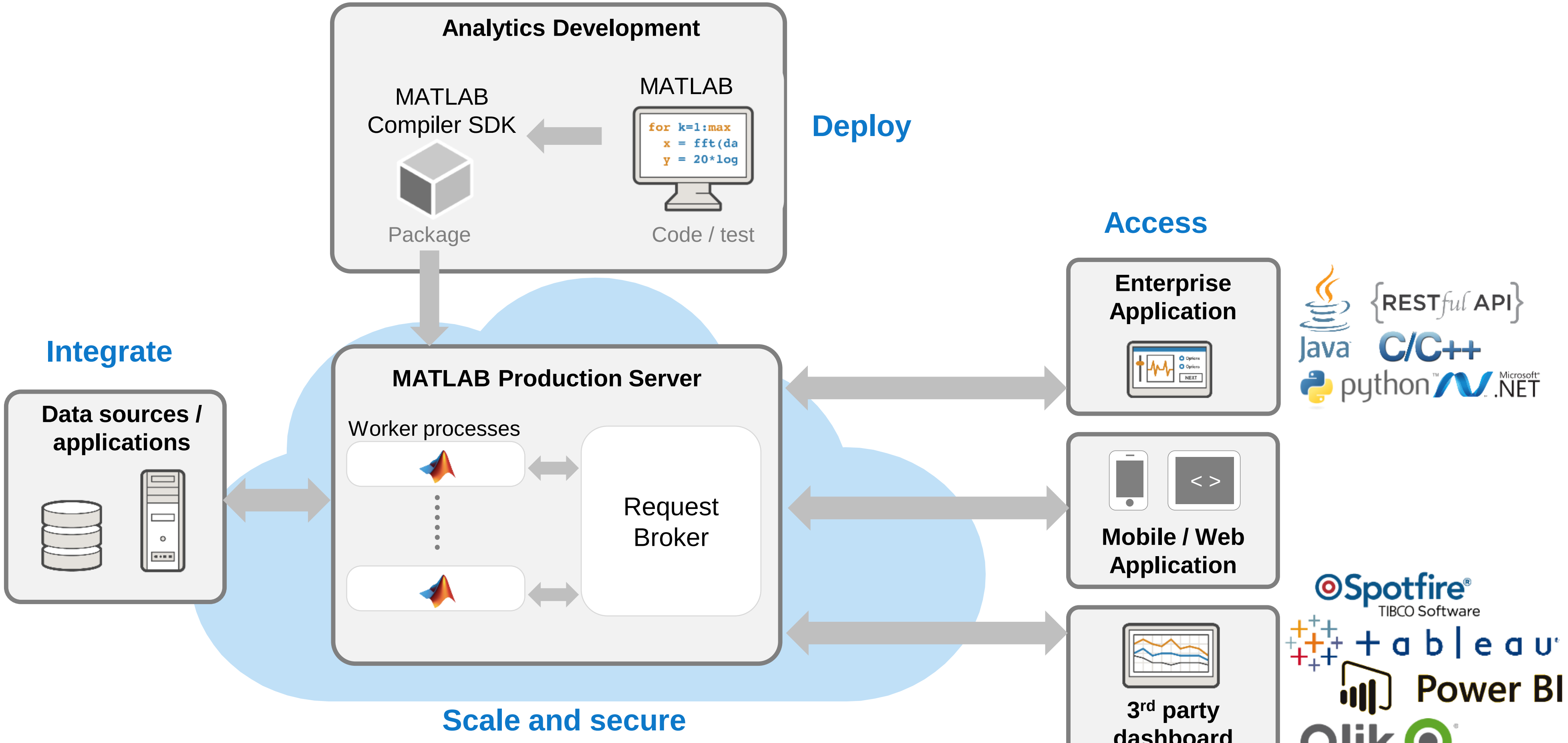
Intel® Xeon® CPU 3.6 GHz - NVIDIA libraries: CUDA9 - cuDNN 7 - Frameworks: TensorFlow 1.8.0, MXNet 1.2.1, PyTorch 0.3.1

Semantic Segmentation

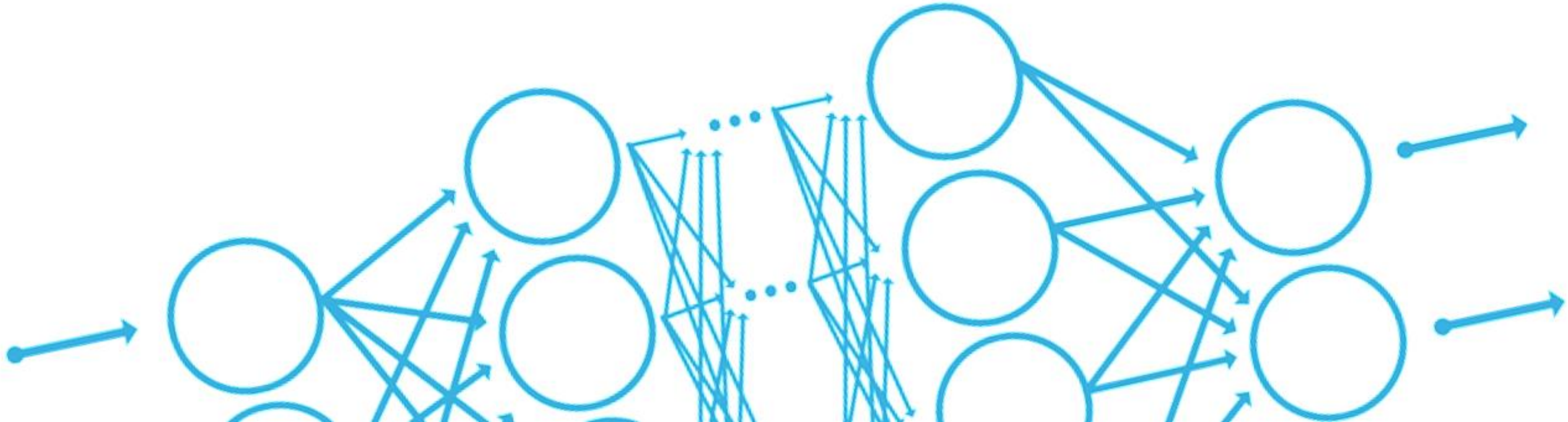
Running in MATLAB

Generated Code from GPU Coder

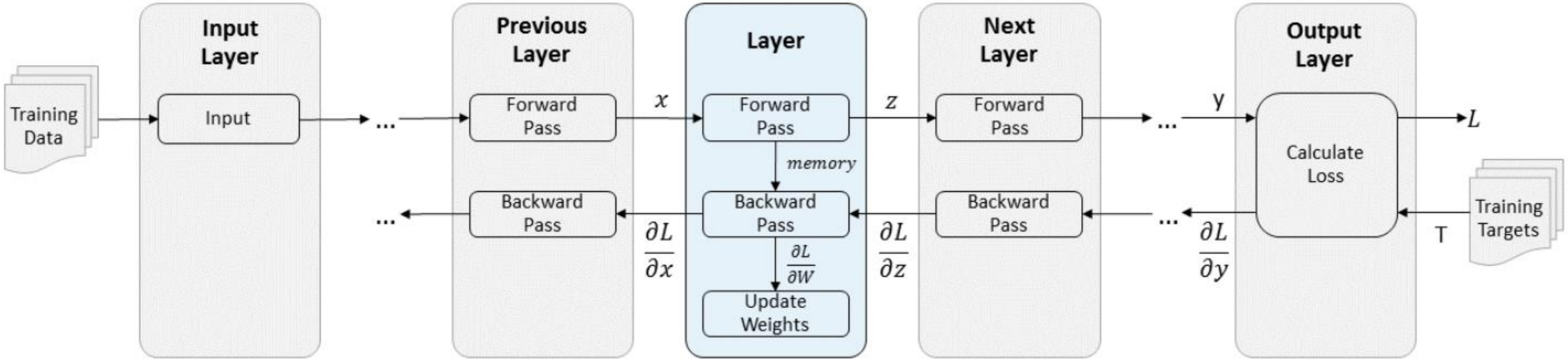
MATLAB Production Server is an application server that publishes MATLAB code as APIs that can be called by other applications

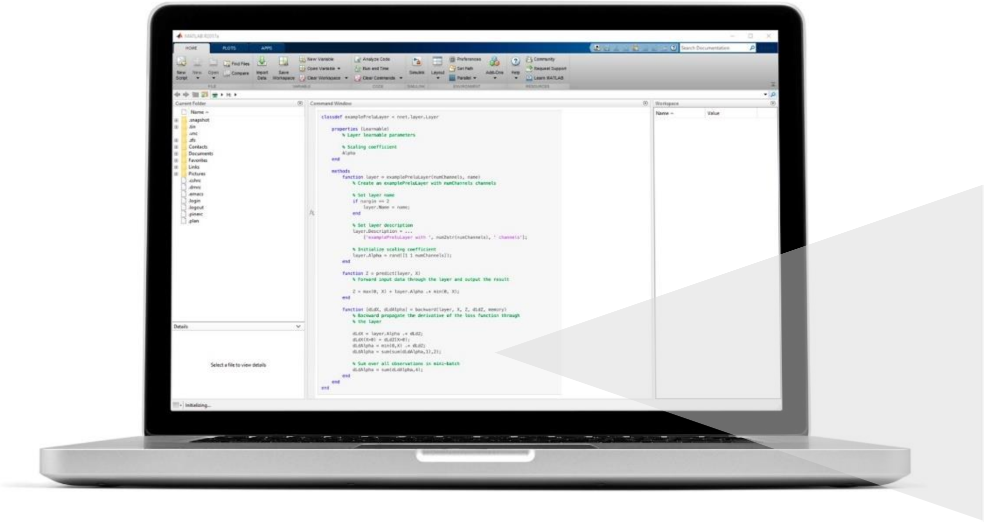


Other Features



Define new deep neural network layers R2017b





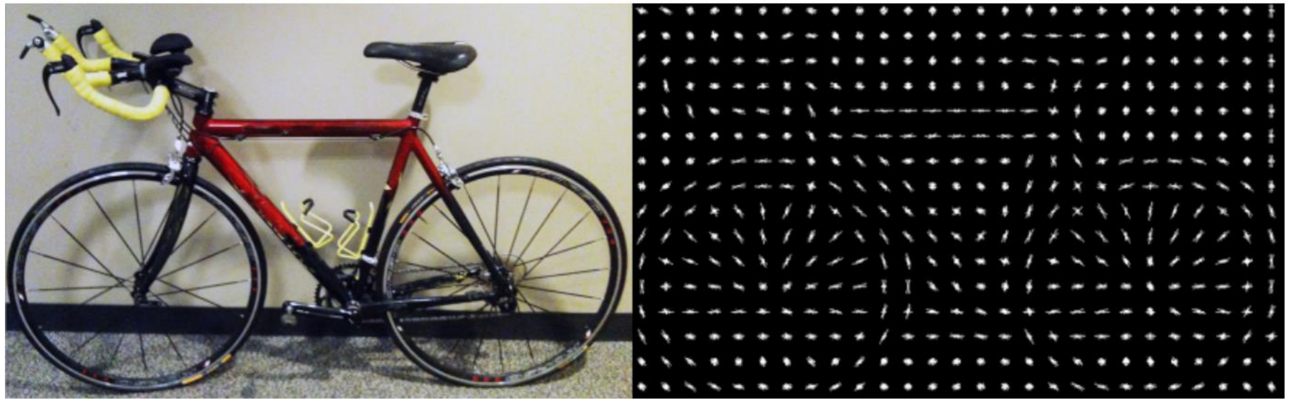
```
function [dLdX, dLdAlpha] = backward(layer, X, Z, dLdZ, memory)
    % Backward propagate the derivative of the loss function through
    % the layer

    dLdX = layer.Alpha .* dLdZ;
    dLdX(X>0) = dLdZ(X>0);
    dLdAlpha = min(0,X) .* dLdZ;
    dLdAlpha = sum(sum(dLdAlpha,1),2);
```

Object Detection Frameworks

R2017a

- Single line of code to train a detector
- Includes:
  - R-CNN
  - Fast R-CNN
  - Faster R-CNN



Create Custom Object Detectors

trainACFObjectDetector	Train ACF object detector
trainCascadeObjectDetector	Train cascade object detector model
trainFastRCNNObjectDetector	Train a Fast R-CNN deep learning object detector
trainFasterRCNNObjectDetector	Train a Faster R-CNN deep learning object detector
trainImageCategoryClassifier	Train an image category classifier
trainRCNNObjectDetector	Train an R-CNN deep learning object detector

Deep learning features overview

- Classification
- Regression *
- Semantic segmentation
- Object detection *
- Scalability *
  - Multiple GPUs
  - Cluster or cloud
- Custom network layers *
- Import models *
  - Caffe
  - Keras/TensorFlow
- Data augmentation *
- Hyperparameter tuning *
  - Bayesian optimization
- Python ⟺ MATLAB interface *
- LSTM networks *
  - Time series, signals, audio
- Custom labeling *
  - API for ground-truth labeling automation
  - Superpixels
- Data validation *
  - Training and testing

** We can cover in more detail outside this presentation*

MATLAB products for deep learning

Required products

- Deep Learning Toolbox
- Parallel Computing Toolbox
- Image Processing Toolbox
- Computer Vision System Toolbox

Optional products

- Statistics and Machine Learning Toolbox
- Signal Processing Toolbox
- Text Analytics Toolbox
- Wavelet Toolbox
- MATLAB Coder
- GPU Coder
- Automated Driving System Toolbox

MathWorks can help you do Deep Learning

Free resources

- **Guided evaluations with a MathWorks deep learning engineer**
- Proof-of-concept projects
- **Deep learning hands-on workshop**
- Seminars and technical deep dives
- Deep learning onramp course

More options

- Consulting services
- Training courses
- Technical support
- Advanced customer support
- Installation, enterprise, and cloud deployment
- Deep Learning Paid Training

Consulting – Deep Learning Discovery

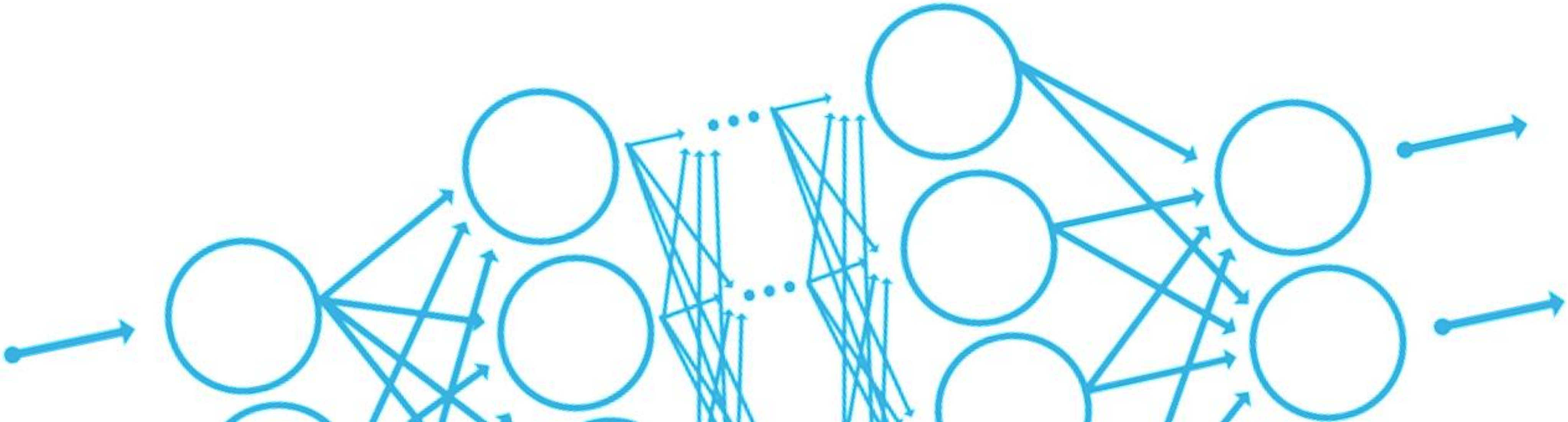
Step 1: Project Assessment

- Proposal for workflow
- Recommendations for methodology & next steps
- Recommendations for tools & licenses
- Tailored slide deck and presentation

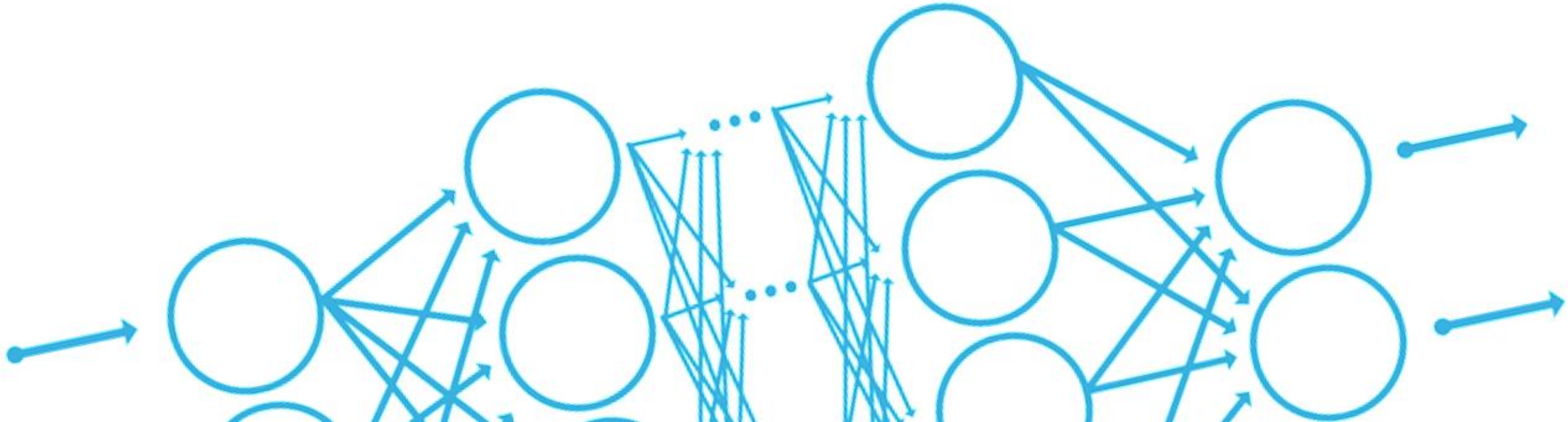
Step 2: Intro to Deep Learning

- Presentation of relevant demos and examples
- Starter code samples
- Guidance regarding further training
- Pointers for additional technical support

Thank you!

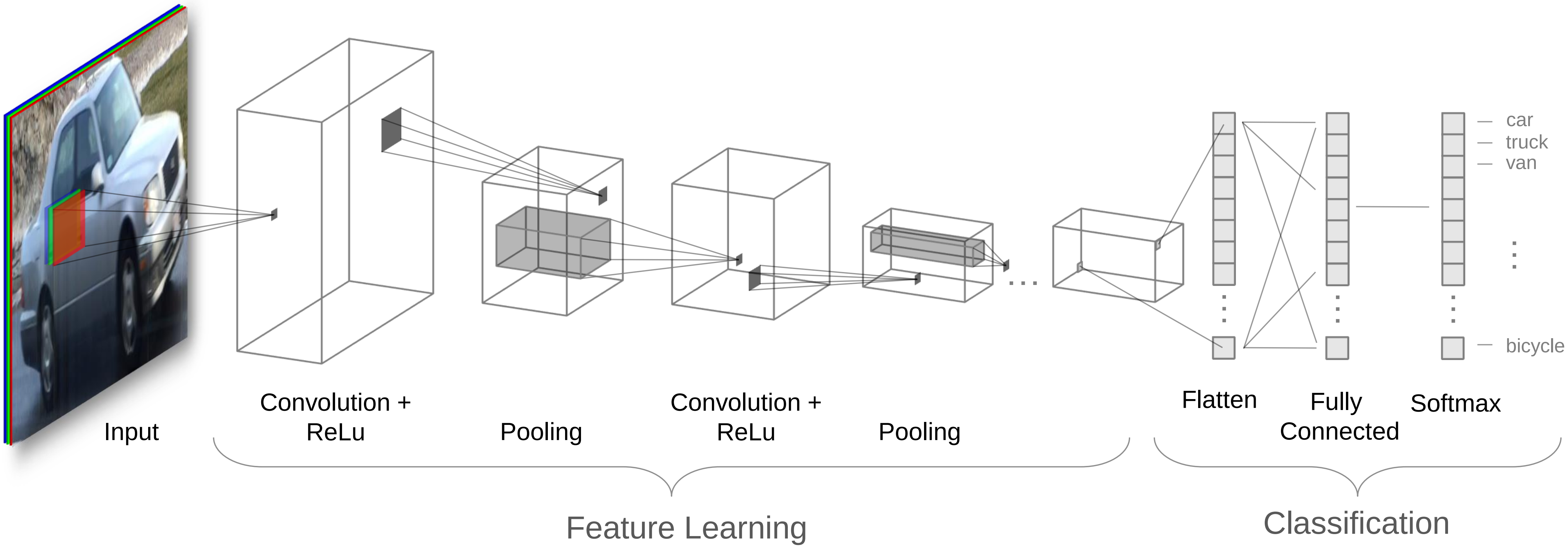


APPENDIX: Extra Slides

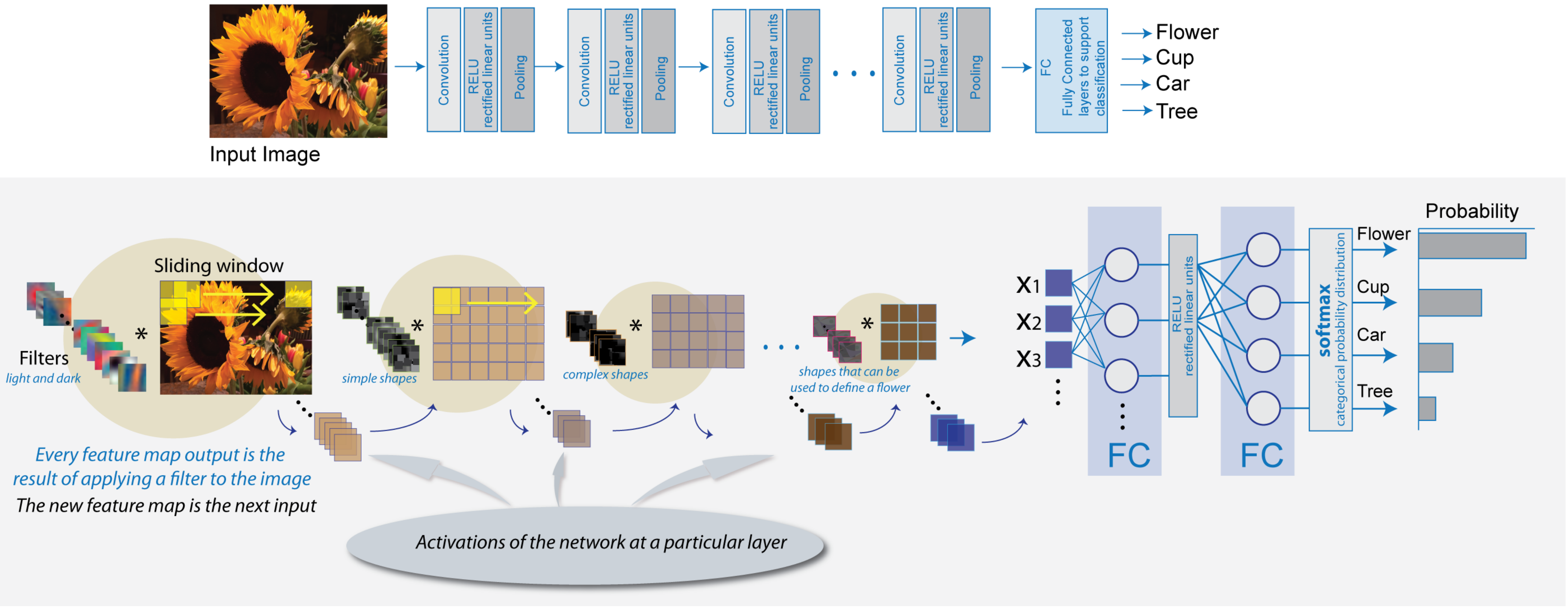


Convolutional Neural Networks

- Train “deep” neural networks on structured data (e.g. images, signals, text)
- Implements Feature Learning: Eliminates need for “hand crafted” features
- Trained using GPUs for performance

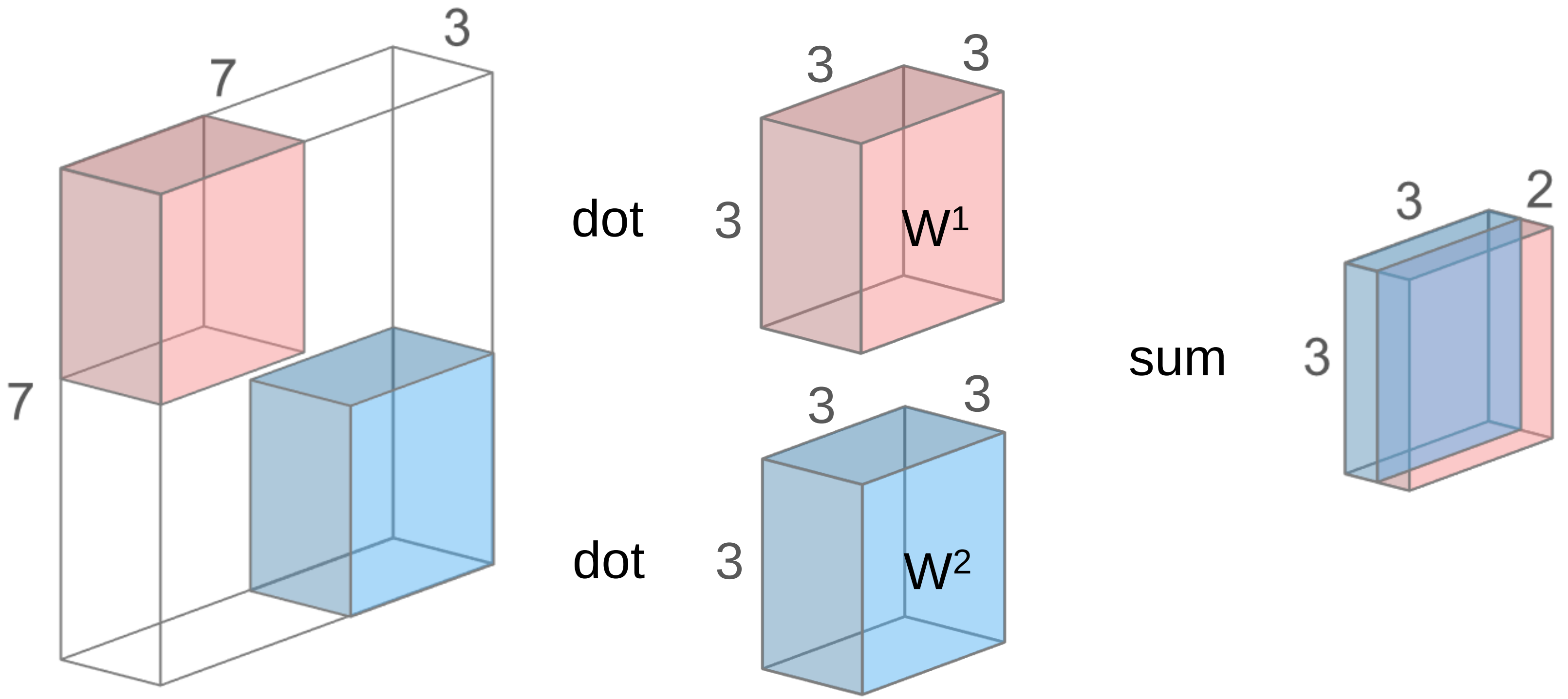


Convolutional Neural Networks



Convolution Layer

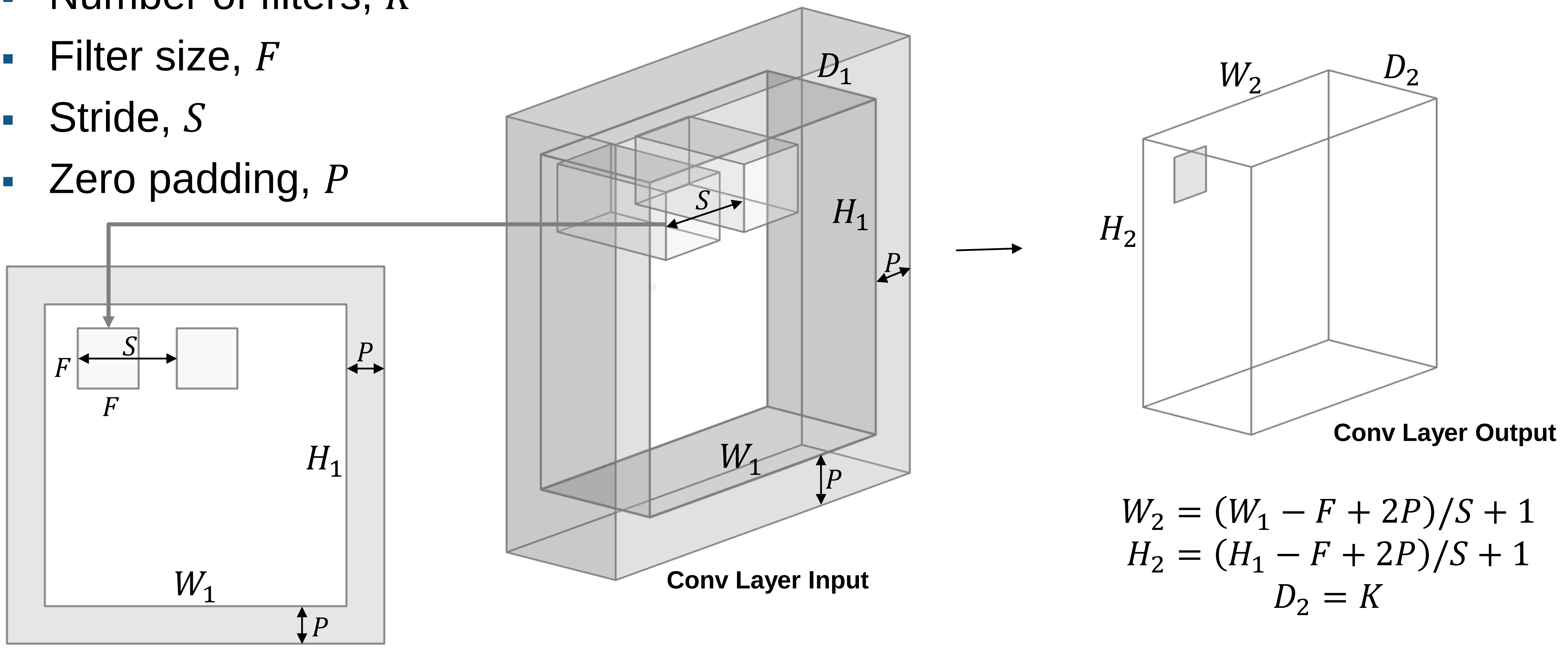
- Core building block of a CNN
- Convolve the filters sliding them across the input, computing the dot product



- Intuition: learn filters that activate when they “see” some specific feature

Convolution Layer – Choosing Hyperparameters

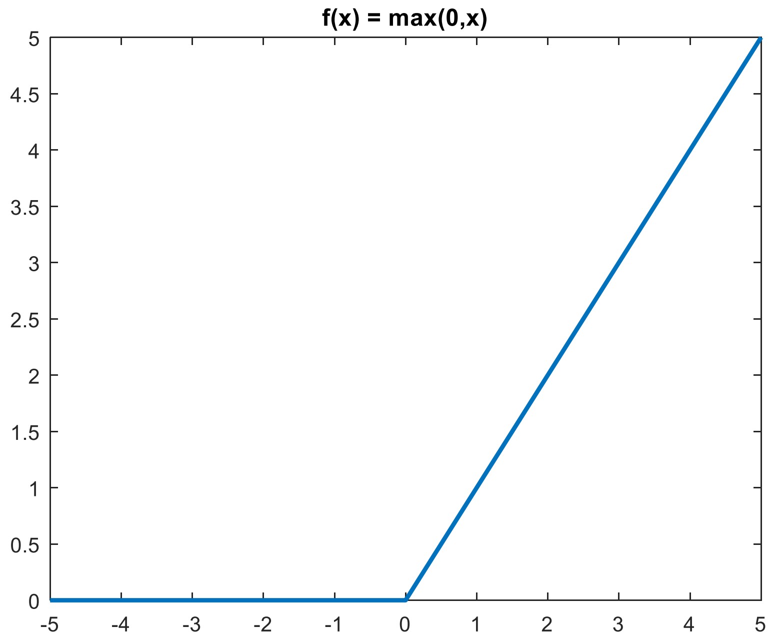
- Number of filters, K
- Filter size, F
- Stride, S
- Zero padding, P



$$W_2 = (W_1 - F + 2P)/S + 1$$
$$H_2 = (H_1 - F + 2P)/S + 1$$
$$D_2 = K$$

Rectified Linear Unit (ReLU) Layer

- Frequently used in combination with Convolution layers
- Do not add complexity to the network
- Most popular choice: $\boldsymbol{f(x) = max(0, x)}$, activation is thresholded at 0



Benefits of Batch Normalization

- Batch normalization reduces the problem of **internal covariate shift**
  – *Internal covariate shift:* Neural networks can be slow to train because of low learning rates and careful parameter optimization – because each layer's inputs change in distribution during training
- It enables higher learning rates
- It regularizes the model

- Source: *Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift* (https://arxiv.org/pdf/1502.03167v3.pdf)

Pooling Layer

- Perform a **downsampling** operation across the spatial dimensions
- Goal: progressively decrease the size of the layers
- Max pooling and average pooling methods
- Popular choice: Max pooling with 2x2 filters, Stride = 2