Agenda

- What’s Deep Learning and why should I care?
- A practical approach to Deep Learning (for images)
  - Transfer the learning from an expert model to your own application
- Building a Deep Learning network from scratch
  - Deep Learning for time series and text data
- Key learnings of the session and cool features
Deep learning with MATLAB is easy and accessible!

Use Less Data and Less Time with Transfer Learning

Automatically generate code!

Prototype and productize your entire workflow on embedded GPUs

Apply Deep Learning not only on images but also on text and time series data
Why Machine Learning or Deep Learning?

- **Transformational technology**
- Close (better) than human accuracy for specific tasks
- **Performance scales with data**
- It is hard to use, it is challenging

Enables engineers, researchers and other domain experts to create products and applications with more built-in intelligence.
Artificial Intelligence, Machine Learning and Deep Learning

Timeline

Artificial Intelligence
- 1950s
- 1997

Machine Learning
- Brain-controlled Robots

Deep Learning
- AlphaGO: 1202 CPUs, 176 GPUs, 100+ Scientists.
- Lee Se-dol: 1 Human Brain, 1 Coffee.

Application Breadth
- 2016
- Today
Machine Learning vs Deep Learning

Machine Learning
- Features
- Training Data
- Learning is an iterative process

Deep Learning
- Deep Learning
- Convolutional Neural Network (CNN)
- End-to-end

Classification Task
- CAR
- TRUCK
- BICYCLE

MathWorks MATLAB EXPO 2018
Deep Learning Common Workflow

**ACCESS AND EXPLORE DATA**
- Files
  - Data Augmentation/Transformation
  - Labeling Automation
  - Import Reference Models

**LABEL AND PREPROCESS DATA**
- Databases
  - Hardware-Accelerated Training
  - Hyperparameter Tuning
  - Network Visualization

**DEVELOP PREDICTIVE MODELS**
- Sensors

**INTEGRATE MODELS WITH SYSTEMS**
- Desktop Apps
  - Desktop Apps
  - Enterprise Scale Systems
  - Embedded Devices and Hardware
Deep learning is usually implemented using a neural network architecture

- The term “deep” refers to the number of layers in the network—the more layers, the deeper the network.

- Data flows through network in layers, which provide transformation of data
Convolutional Neural Network is a popular Deep Learning architecture

<table>
<thead>
<tr>
<th>Shallow Neural Network</th>
<th>Convolutional Neural Network</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Every input neuron connects to every neuron in the hidden layer</th>
<th>Local receptive fields connect to neurons in the hidden layer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Translate across an image to create a feature map efficiently with convolution</td>
<td></td>
</tr>
</tbody>
</table>
# Deep Learning: different types of network architectures

<table>
<thead>
<tr>
<th>Type</th>
<th>Series Network</th>
<th>Directly Acyclic Graph Network (DAG)</th>
<th>Recurrent Neural Network (RNN)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Basic Architecture</strong></td>
<td>Single-in single-out</td>
<td>Multi-in, multi-out No feedback loops</td>
<td>Memory</td>
</tr>
<tr>
<td><strong>Network Example</strong></td>
<td>AlexNet, VGG</td>
<td>R-CNN (fast, faster), GoogLeNet, SegNet</td>
<td>LSTM</td>
</tr>
<tr>
<td><strong>Application</strong></td>
<td>Object Recognition: Cars, Lane, Pedestrian</td>
<td>Object Detection: Cars, Traffic Signs, Scene (Semantic Segmentation)</td>
<td>Sequential data: time series, signals</td>
</tr>
</tbody>
</table>
Agenda

▪ What’s Deep Learning and why should I care?

▪ A practical approach to Deep Learning (for images)
  – Transfer the learning from an expert model to your own application

▪ Building a Deep Learning network from scratch
  – Deep Learning for time series and text data

▪ Key learnings of the session and cool features
Deep Learning can be complex and challenging to apply

Main Challenges
- Handle large image sets
- Image labeling is tedious
- Have access to models

Main Challenges
- Capability of training with multiple GPUs
- Capability of training in the cloud

Main Challenges
- Convert models to CUDA code
- Compress models to fit into embedded GPUs.

Design Deep Learning & Vision Algorithms

Accelerate and Scale Training

High Performance Deployment
Object Recognition using Deep Learning in MATLAB

Supervised Learning

AlexNet
PRETRAINED MODEL

Transfer Knowledge

OfficeNet
NEW MODEL

Stapler
CofeeCoaster
Highlighter
RubikCube

MATLAB Support Package for USB Webcams
Where to start?
Two Approaches for Deep Learning

Train from Scratch

Deep Neural Network, e.g. CNN

LEARNED FEATURES

95%
3%
...
2%

CAR ✓
TRUCK ✗
BICYCLE ✗

Transfer Learning

PRE-TRAINED CNN

Fine - Tune

NEW TASK

CAT ✓
DOG ✗
Why Perform Transfer Learning?

- Leverage best network types from top researchers
- Reference models are great feature extractors
- Less data
- Less training time

Leverage models like AlexNet, VGG-16, and ResNet from top researches.

Models:
- AlexNet
- VGG-16
- ResNet
- Caffe
- GoogLeNet
- TensorFlow/Keras
Transfer Learning Workflow

Load pretrained network

- Early layers that learned low-level features (edges, blobs, colors)
- Last layers that learned task-specific features

1 million images
1000s classes

Replace final layers

New layers learn features specific to your data

Fewer classes
Learn faster
Transfer Learning Workflow

Load pretrained network
- Early layers that learned
  low-level features
  (edges, blobs, colors)
- Last layers that
  learned task
  specific features
- 1 million images
  1000s classes

Replace final layers
- New layers to learn
  features specific
to your data
- Fewer classes
  Learn faster
- 100s images
  10s classes

Train network
- Training images
- Training options
- 100s images
  10s classes
Transfer Learning Workflow

Load pretrained network
- Early layers that learned low-level features (edges, blobs, colors)
- Last layers that learned task-specific features
- 1 million images, 1000s classes

Replace final layers
- Fewer classes
- Learn faster

Train network
- New layers to learn features specific to your data
- 100s images, 10s classes
- Training images, Training options

Predict with trained network
- Deploy results
- Probability
  - Car
  - Truck
  - Bike
Transfer Learning Workflow

Load pretrained network

Early layers that learned low-level features (edges, blobs, colors)

Last layers that learned task specific features

1 million images
1000s classes

Replace final layers

Fewer classes
Learn faster

New layers to learn features specific to your data

Train network

Training images

100s images
10s classes

Train network

Training options

Predict and assess network accuracy

Test images

Trained Network 98 %

Deploy results

Probability
Car
Truck
...
Bike
Accelerate training and prediction!

Learning is an iterative process

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

'ExecutionEnvironment', 'multi-gpu');
```

Alexnet vs Squeezenet

**Alexnet**
- FPS = 471.21
- cash machine: 43.46%
- window shade: 32.75%
- sliding door: 6.43%
- monitor: 4.52%
- mouse: 3.89%

**Squeezenet**
- FPS = 613.20
- web site: 60.41%
- desktop computer: 22.82%
- screen: 7.35%
- monitor: 3.79%
- mouse: 1.60%

More GPUs
Alexnet Inference on NVIDIA Titan Xp

<table>
<thead>
<tr>
<th>Testing platform</th>
<th>CPU</th>
<th>GPU</th>
<th>cuDNN</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Intel(R) Xeon(R) CPU E5-1650 v4 @ 3.60GHz</td>
<td>Pascal Titan Xp</td>
<td>v7</td>
</tr>
</tbody>
</table>

**GPU Coder + TensorRT (3.0.1, int8)**

**GPU Coder + TensorRT (3.0.1)**

**GPU Coder + cuDNN**

**mxNet (1.1.0)**

**TensorFlow (1.6.0)**
Deep Learning is easy and accessible with MATLAB!

**Design Deep Learning & Vision Algorithms**

- **Datastores** for large image sets
- **Automate** image labeling
- **Direct access** to models within MATLAB with support packages
- **Import** Tensor Flow Keras and Caffe networks

**Accelerate and Scale Training**

- **Single line of code** to:
  - **Accelerate** training with multiple GPUs or
  - **Scale** to clusters

**High Performance Deployment**

- **Automate compilation** with GPU Coder
- **1.4x speedup** over C++ Caffe on Jetson TX2
Agenda

- What’s Deep Learning and why should I care?

- A practical approach to Deep Learning (for images)
  - Transfer the learning from an expert model to your own application

- Building a Deep Learning network from scratch
  - Deep Learning for time series and text data

- Key learnings of the session and cool features
Deep Learning for Time Series, Sequences and Text

- Sentiment Analysis
- Speech-to-text
- Text-to-Speech
- Movie Ranking Prediction
- Music Recognition
- Predictive Maintenance
- Instant Translation
Deep Learning for Time Series
Example: Seizure prediction (time series classification)

Goal: Predict seizures in long-term

Dataset: iEGG time series

Data size: 20GB

Output: Classification before or between seizure

Link for melbourne-university-seizure-prediction
Types of Datasets

- **Numeric Data**
- **Time Series/Text Data**
- **Image Data**

- Long Short Term Memory (LSTM)
- LSTM or CNN
- Convolutional Neural Networks (CNN)
  Directed acyclic graph networks (DAG)
### Deep Learning for Time Series

**CNN: Data for Time series = Pixel for Images**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>-0.5</td>
<td>-1</td>
<td>-2.1</td>
</tr>
<tr>
<td>-1.3</td>
<td>0.8</td>
<td>1.1</td>
<td>-2</td>
</tr>
<tr>
<td>1.2</td>
<td>0</td>
<td>1.2</td>
<td>1.6</td>
</tr>
<tr>
<td>0.8</td>
<td>0.7</td>
<td>-0.2</td>
<td>-0.4</td>
</tr>
</tbody>
</table>

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>8</td>
<td>3</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>
Deep Learning for Time Series
Long Short Term Memory (LSTM) Network
Media reported two trees blown down along I-40 in the Old Fort area.

<table>
<thead>
<tr>
<th></th>
<th>cat</th>
<th>dog</th>
<th>run</th>
<th>two</th>
</tr>
</thead>
<tbody>
<tr>
<td>doc1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>doc2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

- Word Docs
- PDF’s
- Text Files
- Stop Words
- Stemming
- Tokenization
- Bag of Words
- TF-IDF
- Word Embeddings
- LSTM
- Latent Dirichlet Allocation
- Latent semantic analysis
Key Takeaways: Deep Learning for Time Series and Text

Applications
- Time series
  - forecasting
  - classification (Predictive Maintenance)
- Text
  - classification (Sentiment Analysis, Tagging)
  - clustering (Topic Modelling)

Text Analytics: Prepend
- Text preprocessing
- Conversion to numeric
Agenda

- What’s Deep Learning and why should I care?
- A practical approach to Deep Learning (for images)
  - Transfer the learning from an expert model to your own application
- Building a Deep Learning network from scratch
  - Deep Learning for time series and text data
- Key learnings of the session and cool features
Deep learning with MATLAB is easy and accessible!

Use Less Data and Less Time with Transfer Learning

Automatically generate code!

Prototype and productize your entire workflow on embedded GPUs

Apply Deep Learning not only on images but also on text and time series data
Algorithm Development, Testing & Verification

Deep Learning
Plot and analyze your network using network analyzer, generate CUDA code that integrates with TensorRT, and deploy deep learning networks to Intel and ARM processors.

Images

Semantic Scene Segmentation

14x with MEX

Acceleration (Classification)

2x with MEX

Deployment (Embedded)

~66 Fps (Tegra X1)
Mini-batchable datastore

Have a small number of large high-res images

To train, need a large number of pairs of images
Ground truth labelling: attributes and sublabels
Visualize and understand the network architecture

Detect problems before wasting time training!

- Missing or disconnected layers,
- Mismatching or incorrect sizes of layer inputs,
- Incorrect number of layer inputs,
- Invalid graph structures.
Semantic Segmentation

- Fully convolutional networks (FCN)
- Segmentation Networks (SegNet)
- Other directed acyclic graph (DAG)

Manage connections, add and remove layers
Manage label data and evaluate performance

>>help semanticseg
Algorithm Development, Testing & Verification
Applicable to other sensors, e.g. LiDar

Access Data
- Velodyne FileReader
- Preprocessing
- Clustering

Preprocess and Label
- LiDAR Annotation
- Ground Truth
- [Under Request]

Object Classification
- PointNet
- Deep Neural Network

Automated Driving System Toolbox™
Computer Vision System Toolbox™
Neural Network Toolbox™
LSTM for both Classification and Regression

Feature extraction

Classification

Time to failure
Define new operations for deep networks with ‘Custom Layers’

```matlab
classdef myLayer < nnet.layer.Layer
    properties (Learnable)
        % Layer learnable parameters go here
    end
    methods
        Z = predict(layer, X)
            % Forward input data through the layer at prediction time and
            % output the result
        [dLdX, dLdW] = backward(layer, X, Z, dLdZ, memory)
            % Backward propagate the derivative of the loss function through
            % the layer
    end
end
```
THANK YOU!

Use Less Data and Less Time with Transfer Learning

Automatically generate code!

Prototype and productize your entire workflow on embedded GPUs

Apply Deep Learning not only on images but also on time series data