MATLAB EXPO 2018
Demystifying Deep Learning
Emelie Andersson
Deep Learning Demo
Image Classification
What is Deep Learning?
Deep Learning

Model learns to perform classification tasks directly from data.
Data Types for Deep Learning

- Signal
- Text
- Image
Deep Learning is Versatile

Detection of cars and road in autonomous driving systems

Rain Detection and Removal

Iris Recognition – 99.4% accuracy

2. Source: An experimental study of deep convolutional features for iris recognition Signal Processing in Medicine and Biology Symposium (SPMB), 2016 IEEE Shervin Minaee; Amirali Abdolrashidi; Yao Wang; An experimental study of deep convolutional features for iris recognition
How is deep learning performing so well?
Deep Learning Uses a Neural Network Architecture

Input Layer

Hidden Layers (n)

Output Layer
Thinking about Layers

- Layers are like blocks
  - Stack them on top of each other
  - Replace one block with a different one

- Each hidden layer processes the information from the previous layer
Thinking about Layers

- Layers are like blocks
  - Stack them on top of each other
  - Replace one block with a different one

- Each hidden layer processes the information from the previous layer

- Layers can be ordered in different ways
1. Read an image to classify
Why MATLAB for Deep Learning?

- MATLAB is Productive
- MATLAB is Fast
- MATLAB Integrates with Open Source
Why MATLAB for Deep Learning?

- MATLAB is Productive
- MATLAB is Fast
- MATLAB integrates with Open Source
“I love to label and preprocess my data”
Caterpillar Case Study

- World’s leading manufacturer of construction and mining equipment.

- Similarity between these projects?
  - Autonomous haul trucks
  - Pedestrian detection
  - Equipment classification
  - Terrain mapping
Computer Must Learn from Lots of Data

- ALL data must first be labeled to create these autonomous systems.

“We were spending way too much time ground-truthing [the data]”
--Larry Mianzo, Caterpillar
How Did Caterpillar Do with Our Tools?

- Semi-automated labeling process

- Used MATLAB for entire development workflow.
  - “Because everything is in MATLAB, development time is short”
Speed up labeling with Image Labeler App
To label an ROI, you must first define one or more of the following label types:
- Rectangle label
- Pixel label

To label a scene, you must first define a scene label.

Load images to start labeling.
MATLAB is Productive

- Image Labeler App semi-automates labeling workflow

- Bootstrapping
  - Improve automatic labeling by updating algorithm as you label more images correctly.
MATLAB is Fast

Performance

Training

Deployment
What is Training?

Feed labeled data into neural network to create working model
Speech Recognition Example

Audio signal → Spectrogram → Image Classification algorithm
Side note: Another Network for Signals - LSTM

- LSTM = Long Short Term Memory (Networks)
  - Signal, text, time-series data
  - Use previous data to predict new information

- I live in France. I speak ____________.
1. Create Datastore

- Datastore creates reference for data
- Do not have to load in all objects into memory

```matlab
datafolder = fullfile(tempdir,'speech_commands_v0.01');
addpath(fullfile(matlabroot,'toolbox','audio','audiodemos'))
ads = audioexample.Datastore(datafolder, ...'
    'IncludeSubfolders',true, ...'
    'FileExtensions','.wav', ...'
    'LabelSource','foldernames', ...'
    'ReadMethod','File')
```
2. Compute Speech Spectrograms

![Speech Spectrogram Examples]

- **Amplitude**
  - Examples for words "up" and "go".

- **Frequency**
  - Spectrogram for words "up" and "go".

- **Time**
  - Time axis for spectrogram examples.
3. Split datastores

**Training**
- 70%
- Trains the model
- Computer “learns” from this data

**Validation**
- 15%
- Checks accuracy of model during training

**Test**
- 15%
- Tests model accuracy
- Not used until validation accuracy is good
4. Define Architecture and Parameters

Neural Network Architecture

```matlab
layers = [
    imageInputLayer(ImageSize)
    convolution2dLayer(3,16,'Padding','same')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2)
    convolution2dLayer(3,32,'Padding','same')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer(2,'Stride',2,'Padding',[0,1])
    dropoutLayer(dropoutProb)
    convolution2dLayer(3,64,'Padding','same')
    batchNormalizationLayer
    reluLayer
    dropoutLayer(dropoutProb)
    convolution2dLayer(3,64,'Padding','same')
    batchNormalizationLayer
    reluLayer
    maxPooling2dLayer([1 13])
    fullyConnectedLayer(numClasses)
    softmaxLayer
    weightedCrossEntropyLayer(classNames, classWeights)];
```

Model Parameters

```matlab
miniBatchSize = 128;
validationFrequency = floor(numel(YTrain)/miniBatchSize);
options = trainingOptions('adam', ...
    'InitialLearnRate',5e-4, ...
    'MaxEpochs',25, ...
    'MiniBatchSize',miniBatchSize, ...
    'Shuffle','every-epoch', ...
    'Plots','training-progress', ...
    'Verbose',false, ...
    'ValidationData',{XValidation,YValidation}, ...
    'ValidationFrequency',validationFrequency, ...
    'ValidationPatience',Inf, ...
    'LearnRateSchedule','piecewise', ...
    'LearnRateDropFactor',0.1, ...
    'LearnRateDropPeriod',20);
```
5. Train Network
Deep Learning on CPU, GPU, Multi-GPU and Clusters

**How to Target?**

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

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

```matlab
opts = trainingOptions('sgdm', ...
    'MaxEpochs', 100, ...
    'MiniBatchSize', 250, ...
    'InitialLearnRate', 0.00005, ...
    'ExecutionEnvironment', 'parallel');
```
MATLAB is Fast for Deployment

- Target a GPU for optimal performance
- NVIDIA GPUs use CUDA code
- We only have MATLAB code. Can we translate this?
GPU Coder

- Automatically generates **CUDA** Code from MATLAB Code
  - can be used on NVIDIA GPUs
Prediction Performance: Fast with GPU Coder

Why is GPU Coder so fast?
- Analyzes and optimizes network architecture
- Invested 15 years in code generation

Images/Sec

Prediction (V100 GPU)

TensorFlow
MATLAB
MXNet
GPU Coder
Why MATLAB?

- MATLAB is Productive
- MATLAB is Fast
- MATLAB Integrates with Open Source
Used MATLAB and Open Source Together

- Used Caffe and MATLAB together
- Use our tools where it makes your workflow easier!

MATLAB Integrates with Open Source Frameworks

- Access to many pretrained models through add-ons
- Users wanted to import latest models

- Import models directly from Tensorflow or Caffe
  - Allows for improved collaboration

KERAS IMPORTER
Importer for TensorFlow-Keras Models

Caffe MODELS
Keras-Tensorflow Importer
MATLAB Integrates with Open Source Frameworks

- MATLAB supports entire deep learning workflow
  - Use when it is convenient for your workflow
- Access to latest models
- Improved collaboration with other users
Why MATLAB for Deep Learning?

- MATLAB is Productive
- MATLAB is Fast (*Performance*)
- MATLAB Integrates with Open Source (*Frameworks*)