MATLAB EXPO 2018

Demystifying Deep Learning

“Let the computers do the hard work”

Jérémy Huard
Why MATLAB for Deep Learning?

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

Model learns to perform 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
Deep Learning in 6 Lines of MATLAB Code

1. Read an image to classify
Why MATLAB for Deep Learning?

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

~ Said no engineer, ever.
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
  - “We go from having to label 100 percent of our data to only having to label about 80 to 90 percent”

- Used MATLAB for entire development workflow.
  - “Because everything is in MATLAB, development time is short”
How Does MATLAB Come into Play?
Labeling Videos: Ground Truth Labeler App
Labeling pixels
MATLAB is Productive

- Image Labeler App semi-automates labeling workflow
  
- Improve automatic labeling by updating algorithm as you label more images correctly.
  
- Easy to load metadata even when labeling manually
MATLAB is Fast

Performance

Training

Deployment
Speech Recognition Example

Audio signal → Spectrogram → Image Classification algorithm
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
2. Compute Speech Spectrograms

- **Amplitude**
  - "up" sound: Amplitude changes with time.
  - "go" sound: Amplitude changes with time.
  - "up" sound: Amplitude changes with time.

- **Frequency**
  - "up" sound: Frequency bands across time.
  - "go" sound: Frequency bands across time.
  - "up" sound: Frequency bands across time.

- **Time**
  - Time axis for all graphs.
  - Time axis for all graphs.
  - Time axis for all graphs.
### 3. Split datastores

<table>
<thead>
<tr>
<th></th>
<th>Training</th>
<th>Validation</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>70%</td>
<td>15%</td>
<td>15%</td>
</tr>
</tbody>
</table>

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

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

- **Test**
  - 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 
    maxPooling2dLayer(2,'Stride',2,'Padding',[0,1])
    dropoutLayer(dropoutProb)
    convolution2dLayer(3,64,'Padding','same') 
    batchNormalizationLayer 
    reluLayer 
    maxPooling2dLayer(2,'Stride',2,'Padding',[0,1])
    dropoutLayer(dropoutProb)
    convolution2dLayer(3,64,'Padding','same') 
    batchNormalizationLayer 
    reluLayer 
    maxPooling2dLayer(2,'Stride',2,'Padding',[0,1])
    dropoutLayer(dropoutProb)
    convolution2dLayer(3,13,'Padding','same') 
    batchNormalizationLayer 
    reluLayer 
    fullyConnectedLayer(numClasses) 
    softmaxLayer 
    weightedCrossEntropyLayer(classNames,classWeights)];
```

Training Parameters

- `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);`
4. How to choose the right structure to start from?

**Pretrained Models for Transfer learning**
- Access pretrained models from top researchers with a single line of code
  ```
  net = alexnet;
  net = vgg16;
  net = inceptionv3;
  net = googlenet;
  ...
  ```

**Import Models from OSS Frameworks**
- Caffe Model Importer
- TensorFlow/Keras Model Importer

**Start from scratch**
- Verify structure with Network Analyzer App
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' );
```
Training Performance

Seconds / Epoch

Training (V100 GPU) Lower is better

TensorFlow
MATLAB
MXNet
Batch size 32
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

- CUDA extends C/C++ code with constructs for parallel computing
GPU Coder Performance

Inference with MATLAB
Why MATLAB?

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

- Used Caffe and MATLAB together
- Achieved significantly better results than an engineered rain model.
- 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
- MATLAB Integrates with Open Source
MathWorks Courses

MATLAB Courses

Learn MATLAB for Free
Hands-on practice sessions and demonstrations
Launch MATLAB Onramp
View my courses

Get Started

MATLAB Onramp
Get started quickly with the basics of MATLAB.
Launch

Deep Learning Onramp
Get started quickly using deep learning methods to perform image recognition.
Launch

Core MATLAB Functionality