MATLAB EXPO 2018

Demystifying Deep Learning

“Let the computers do the hard work”

Giuseppe Ridinò
Deep Learning Demo
Image Classification
Why MATLAB for Deep Learning?

- MATLAB is Productive
- MATLAB is Fast
- MATLAB Integrates with Open Source
Deep Learning Applications

Voice assistants (speech to text)
Teaching character to beat video game
Automatically coloring black-and-white images
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

Iris Recognition – 99.4% accuracy

Rain Detection and Removal

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 Abdorashidiy; 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 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
“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
What 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?
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
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

Amplitude

Frequency

Time
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

Model Parameters

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)];

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
Training is an Iterative Process

```
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);
```

Parameters adjusted according to performance
Deep Learning on CPU, GPU, 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

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

Inference with MATLAB
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

<table>
<thead>
<tr>
<th>Model</th>
<th>Images/Sec</th>
</tr>
</thead>
<tbody>
<tr>
<td>AlexNet</td>
<td></td>
</tr>
<tr>
<td>ResNet-50</td>
<td></td>
</tr>
<tr>
<td>VGG-16</td>
<td></td>
</tr>
</tbody>
</table>

TensorFlow
MATLAB
MXNet
GPU Coder

MATLAB EXPO 2018
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 EXPO 2018
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-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 \textit{(Performance)}
- MATLAB Integrates with Open Source \textit{(Frameworks)}