Computer Vision System Design
(Deep Learning and 3D vision)

Mandar Gujrathi
Applications Engineer
Mandar.Gujrathi@mathworks.com.au

Why are we talking about Computer Vision today?

INTEL ACQUIRES COMPUTER VISION FOR IOT, AUTOMOTIVE

By Doug Davis

Computer Vision Hardware and Software Market to Reach $48.6 Billion by 2022, According to Tractica

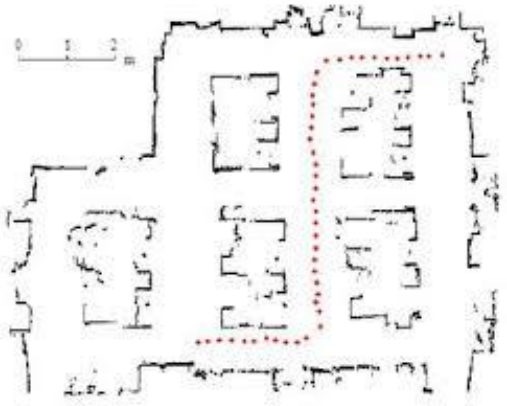
Automotive and Consumer Markets to Surge in the Next Few Years, with Strong Growth Also Expected for Robotics and Security Applications

Human-in-the-loop deep learning will help drive autonomous cars

NAVEEN RAO, NERVANA SYSTEMS JUNE 25, 2016 11:30 AM

Agenda

- Stereo Vision



- Deep Learning

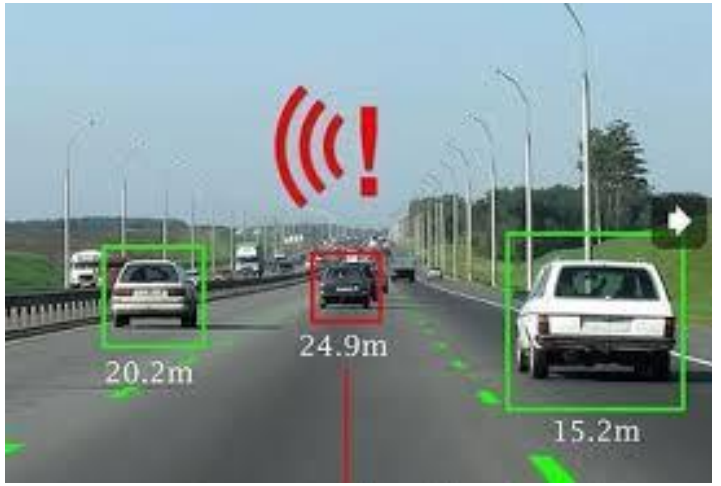
Stereo Vision

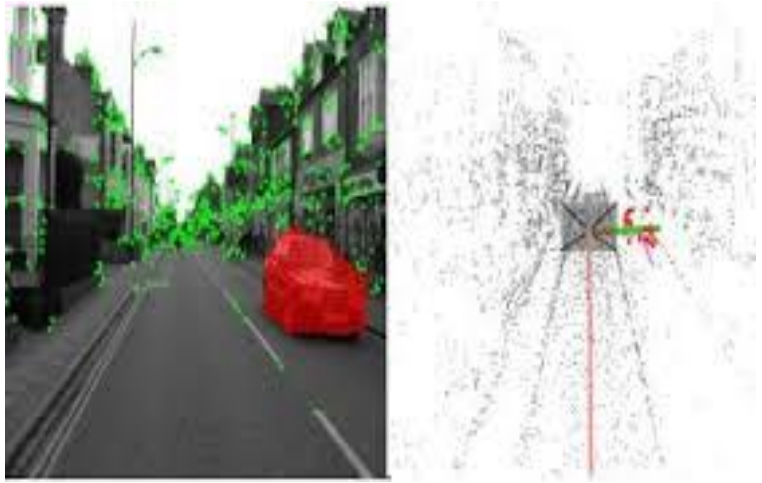
- *Stereo vision is the process of extracting 3-D information from multiple 2-D views of a scene.*

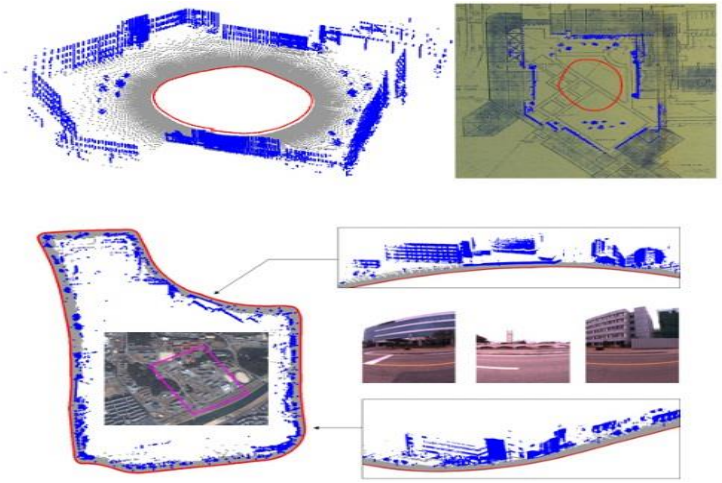
- *Where can I find this capability?*

Computer Vision System Toolbox

Point Cloud Application : ADAS



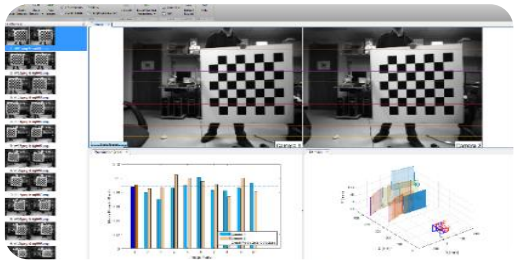




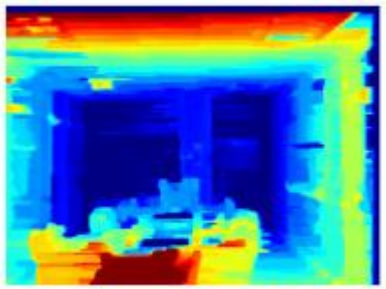
[MathWorks ADAS e-book]

Stereo vision workflow

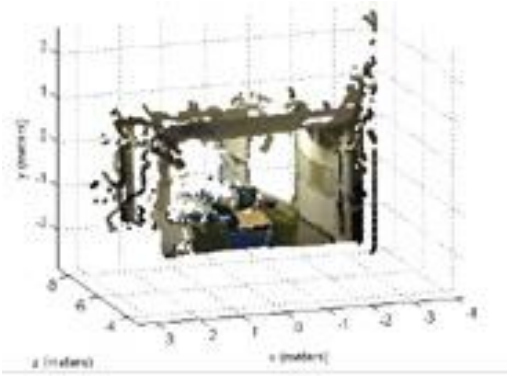
Image Rectification
RectifyStereoImages()



Perform Stereo Camera Calibration
Get parameters



Compute Pixel Differences:
Disparity map
disparity()



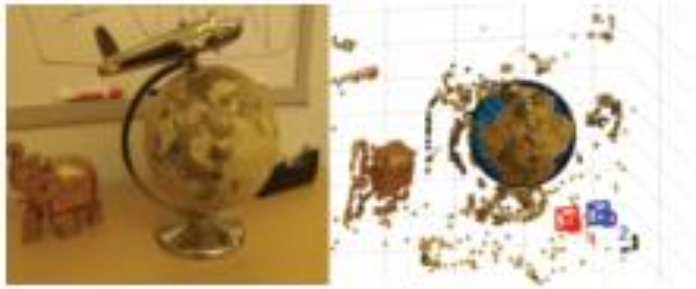
3D point cloud
pcshow()

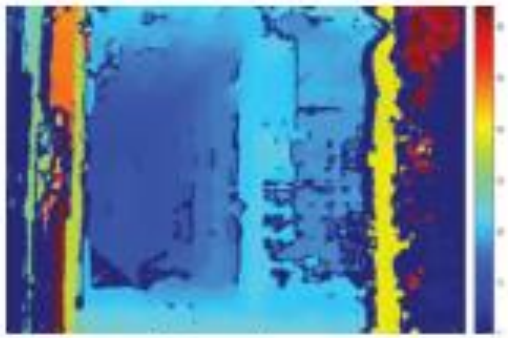
[Stereo Vision]

Camera Calibration

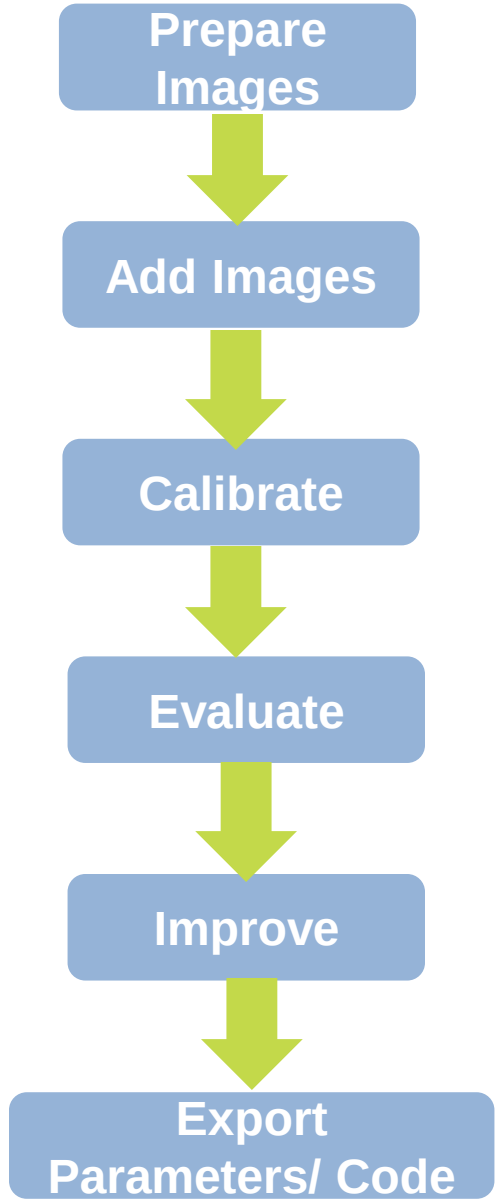
- Estimates the parameters of video camera.

Remove Lens Distortion

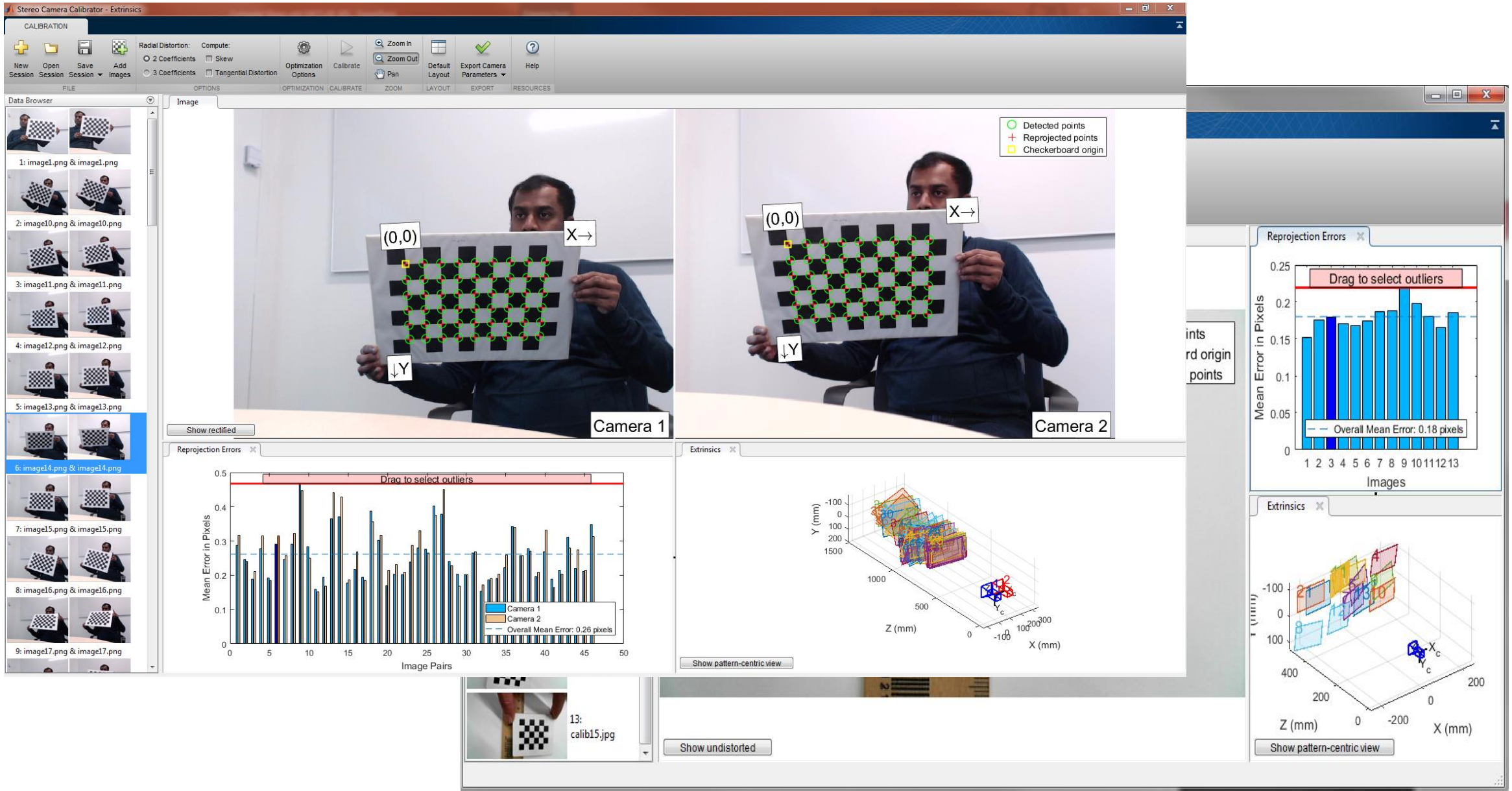

Estimate 3-D Structure from Camera Motion


Estimate Depth
Using a Stereo Camera

Measure Planar Objects

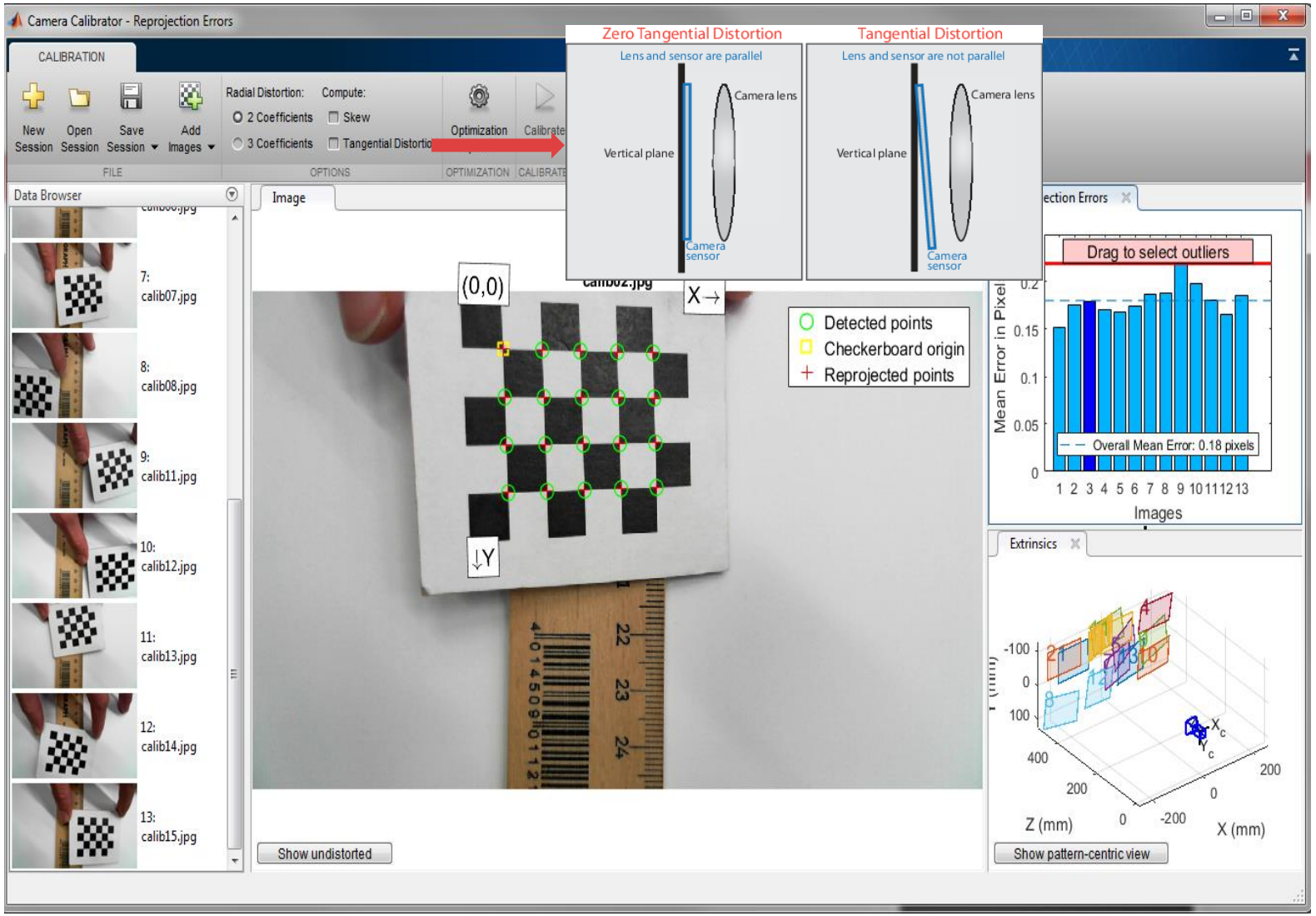


Camera Calibration Apps



Camera Calibration App Advantages

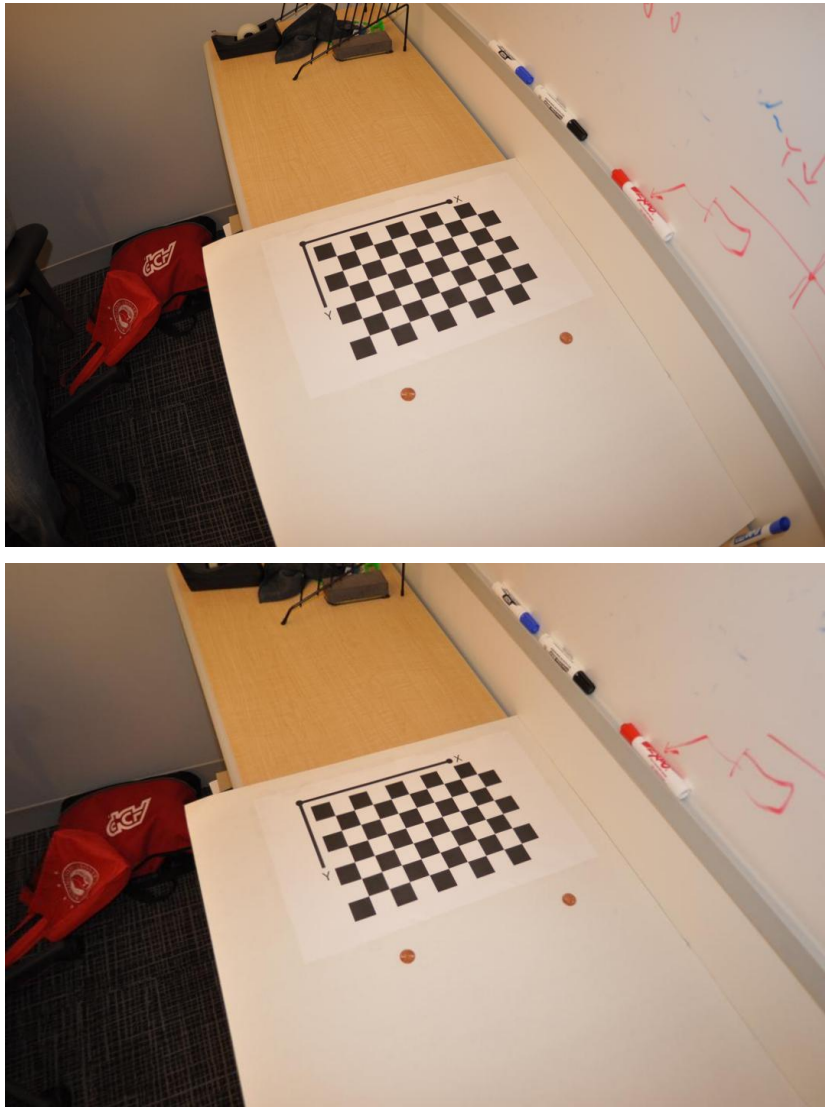
- Simplified workflow for estimating camera intrinsic and extrinsic parameters
- Removes the effects of lens distortion from an image
- Automatically detects checkerboard patterns
- Helps to evaluate accuracy
- Generates code !



Remove Lens Distortion From an Image

Removes radial and tangential distortion.

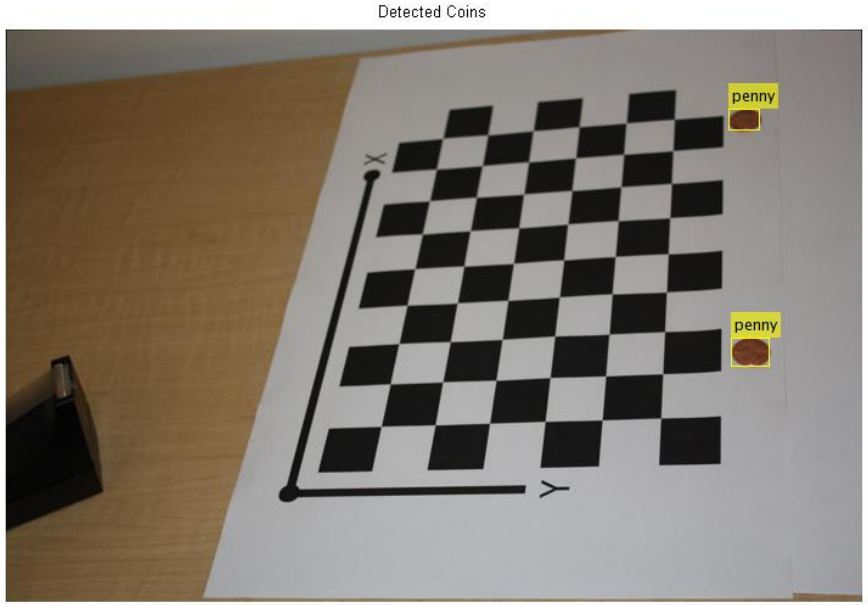
- Radial distortion (“barrel” or “pincushion”) is caused by the curvature of the lens
- Tangential distortion is caused by misalignment between the lens and the sensor



Measuring Planar Objects With a Calibrated Camera

Featured example: measure the diameter of a penny in millimeters.

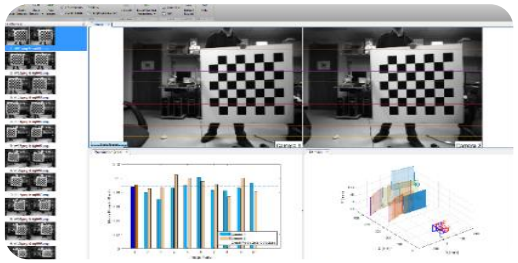
- Undistort the image
- Detect the penny
- Project points from the image into the world
- Measure the diameter in millimeters



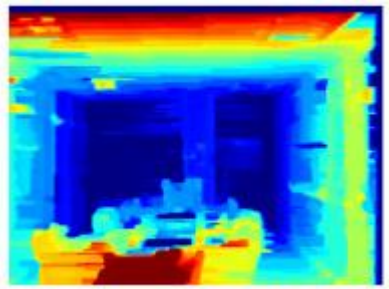

```
>> showdemo('MeasuringPlanarObjectsExample')
```

Stereo vision workflow

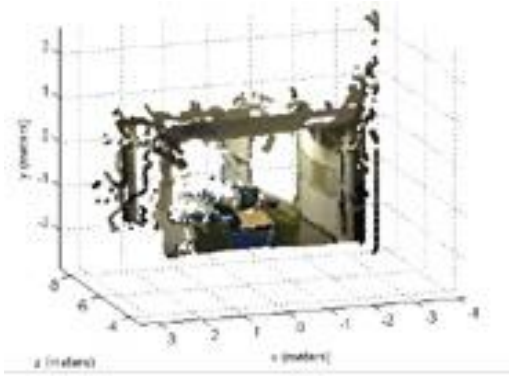
Image Rectification
RectifyStereoImages()



Perform Stereo Camera Calibration
Get parameters



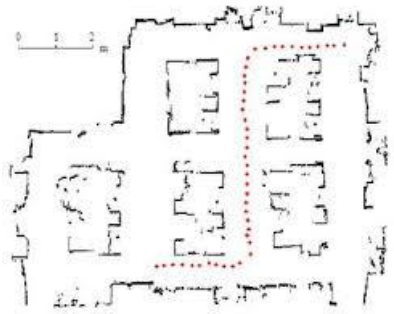
Compute Pixel Differences:
Disparity map
disparity()

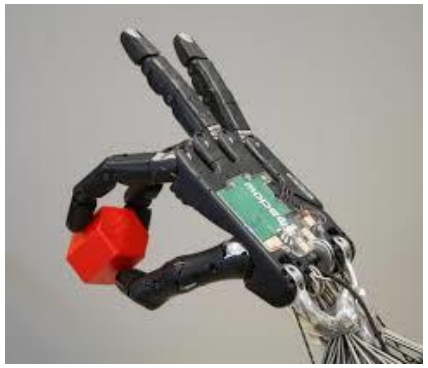


3D point cloud
pcshow()

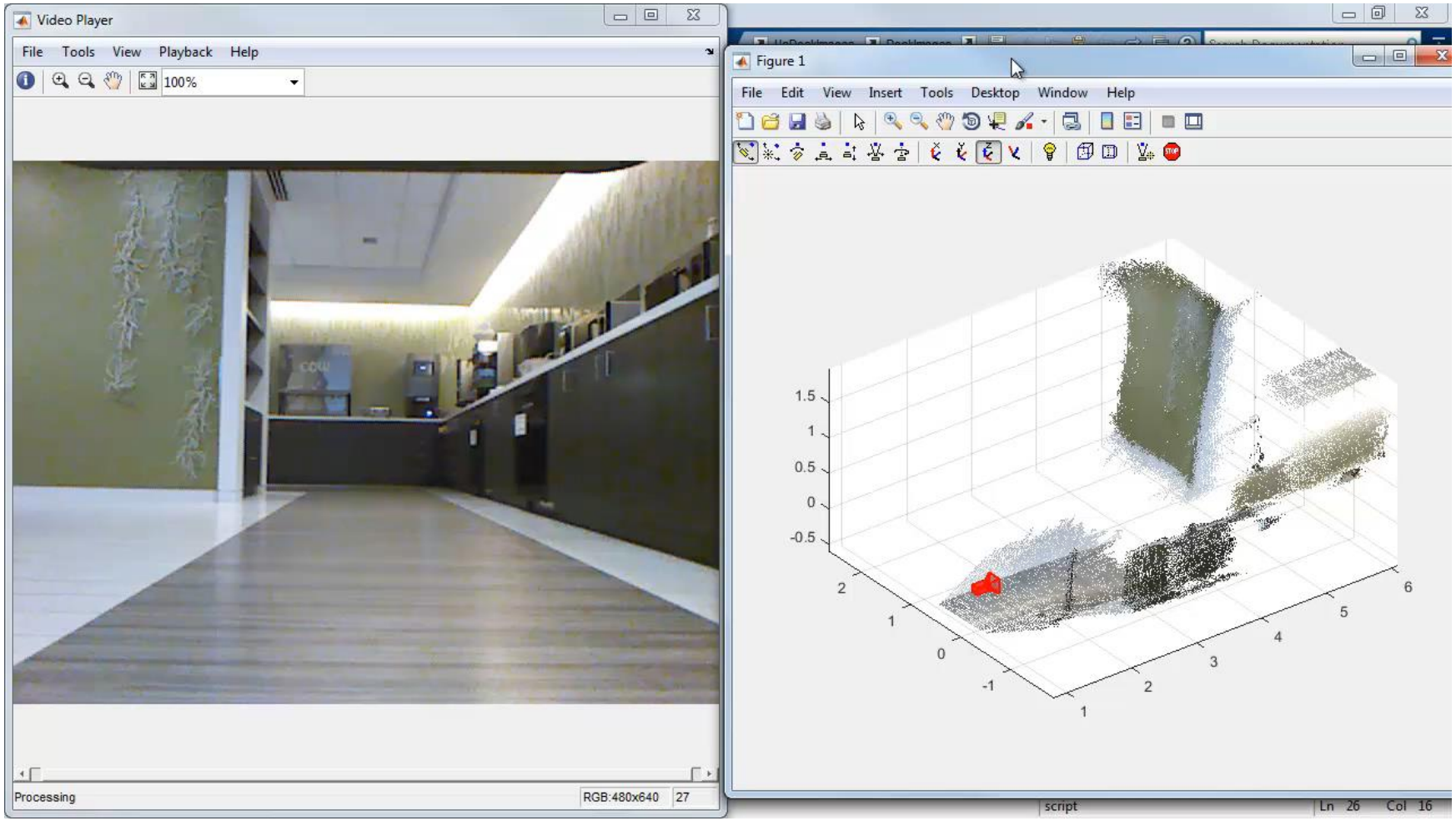
[Stereo Vision]

Point Cloud Application – Robot Vision.

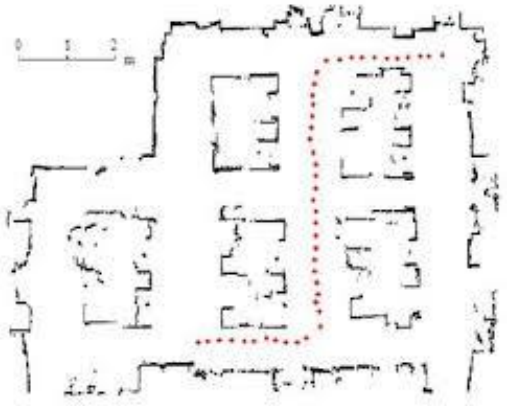




3-D Vision for Robotics



Agenda

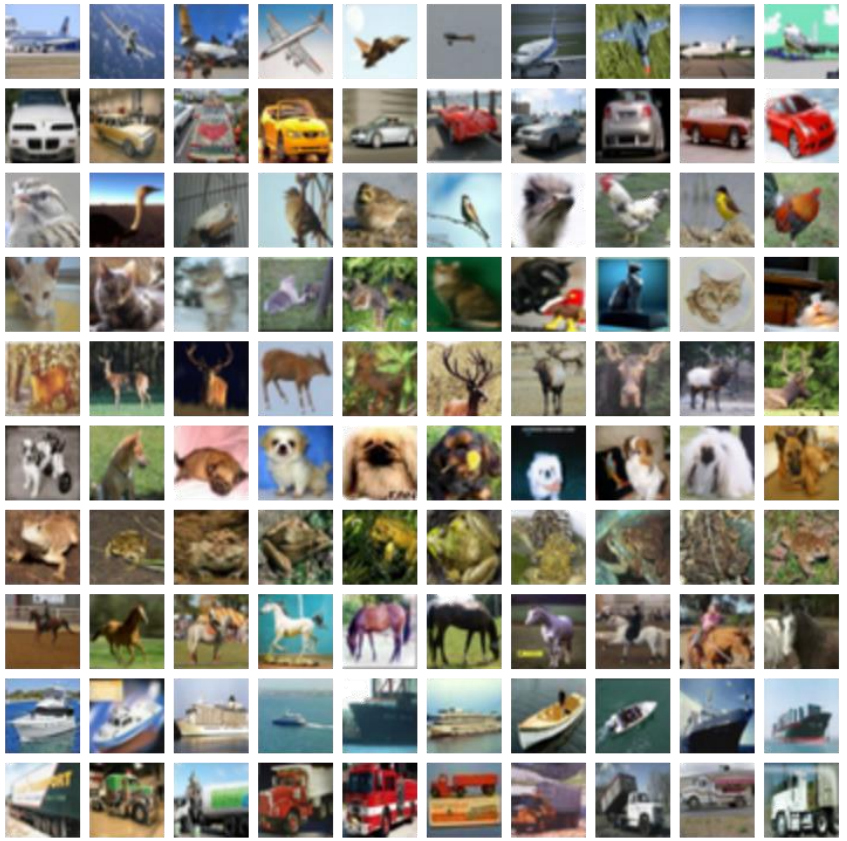


- Stereo Vision
- Deep Learning

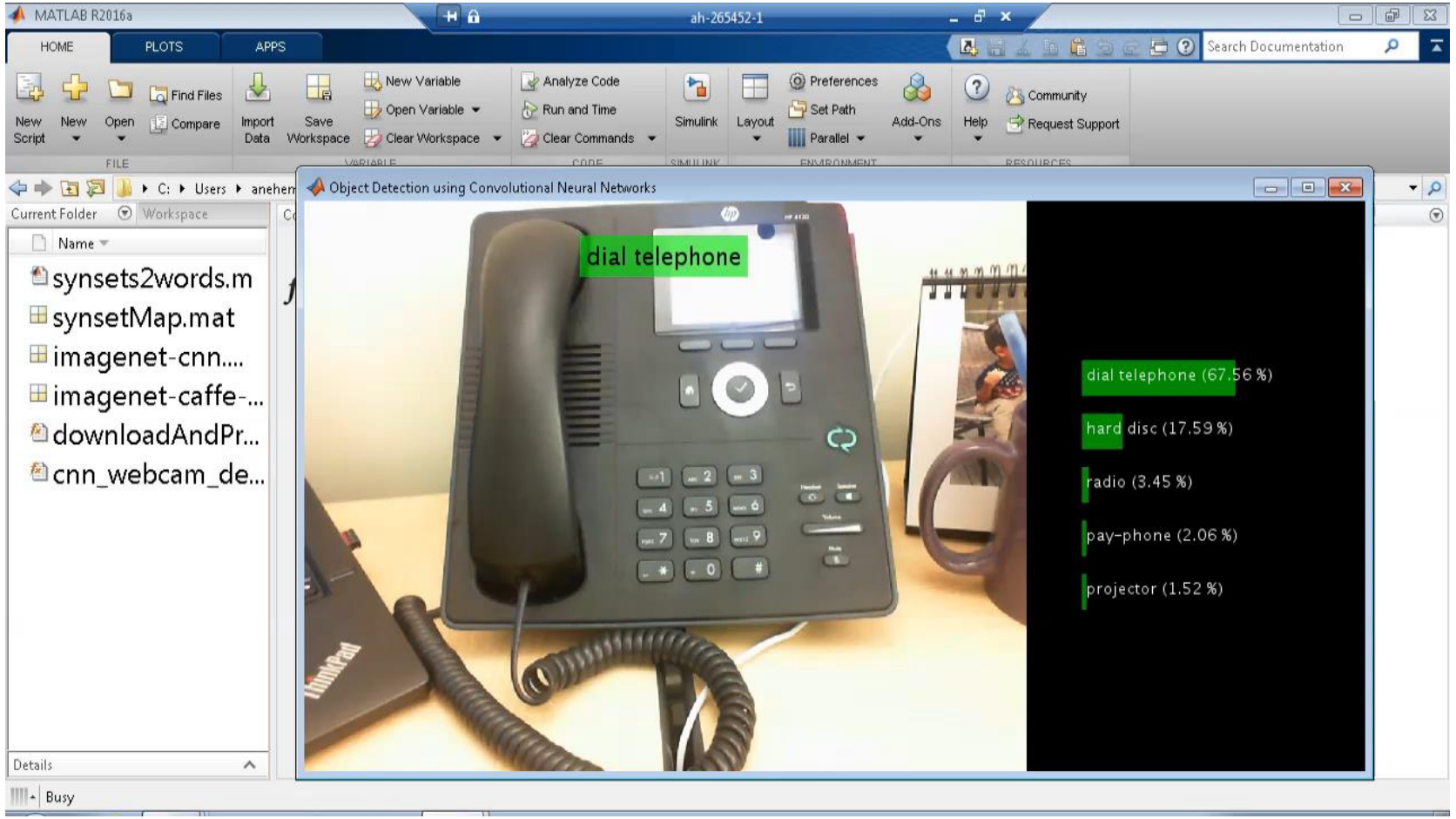
Deep Learning

How long it took us to…

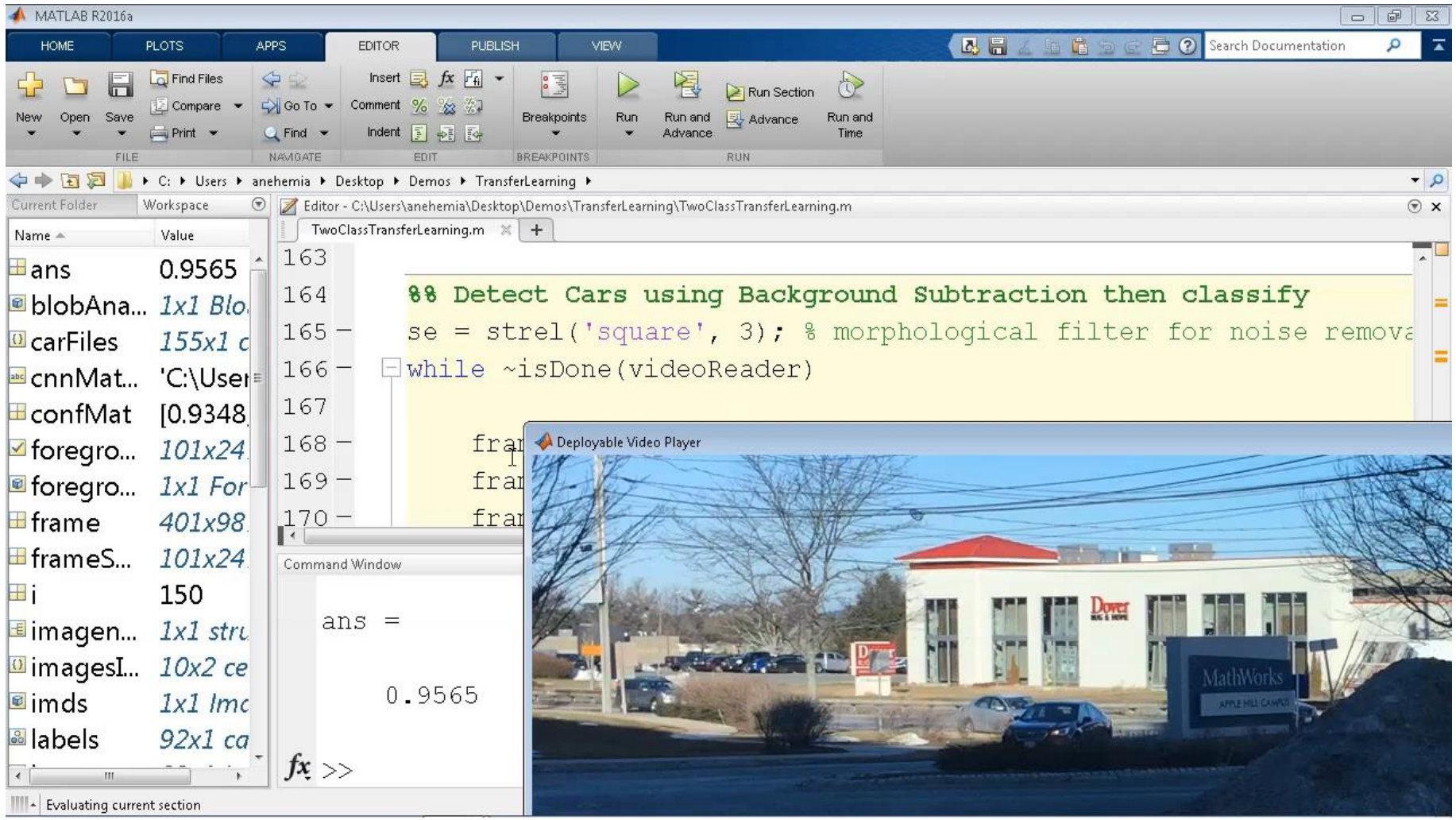
- Question – At what age does a person recognise:
  - Car or Plane
  - Car or SUV
  - Toyota or Mazda



Demo : Live Object Recognition with Webcam



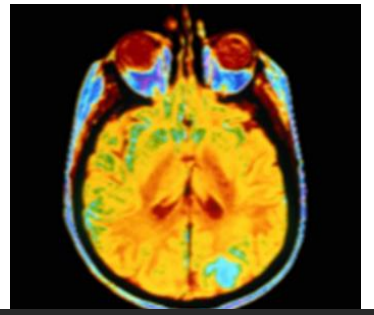
Demo


MATLAB R2016a
HOME
PLOTS
APPS
EDITOR
PUBLISH
VIEW
Search Documentation
New
Open
Save
Find Files
Compare
Print
Go To
Find
Insert
Comment
Indent
Breakpoints
Run
Run and Advance
Run Section
Advance
Run and Time
FILE
NAVIGATE
EDIT
BREAKPOINTS
RUN
C: ▸ Users ▸ anehemia ▸ Desktop ▸ Demos ▸ TransferLearning
Current Folder
Workspace
Name
Value
ans 0.9565
blobAna... 1x1 Blo
carFiles 155x1 c
cnnMat... 'C:\User
confMat [0.9348
foregro... 101x24
foregro... 1x1 For
frame 401x98
frameS... 101x24
i 150
imagen... 1x1 stru
imagesI... 10x2 ce
imds 1x1 Imc
labels 92x1 ca
Editor - C:\Users\anehemia\Desktop\Demos\TransferLearning\TwoClassTransferLearning.m
TwoClassTransferLearning.m
163
164 %% Detect Cars using Background Subtraction then classify
165 se = strel('square', 3); % morphological filter for noise remova
166 while ~isDone(videoReader)
167
168
169
170
Command Window
ans =
0.9565
>>
Deployable Video Player
MathWorks
Evaluating current section

Deep Learning is Ubiquitous

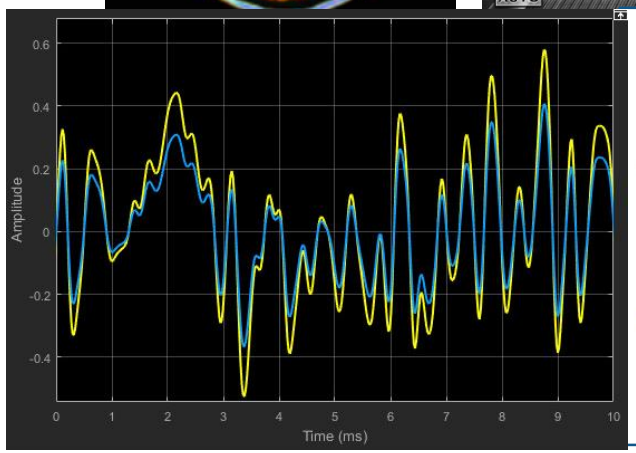
Computer Vision

- Pedestrian and traffic sign detection
- Landmark identification
- Medical diagnosis and drug discovery
- Surveillance



Text and Signal Processing

- Speech Recognition
- Speech & Text Translation

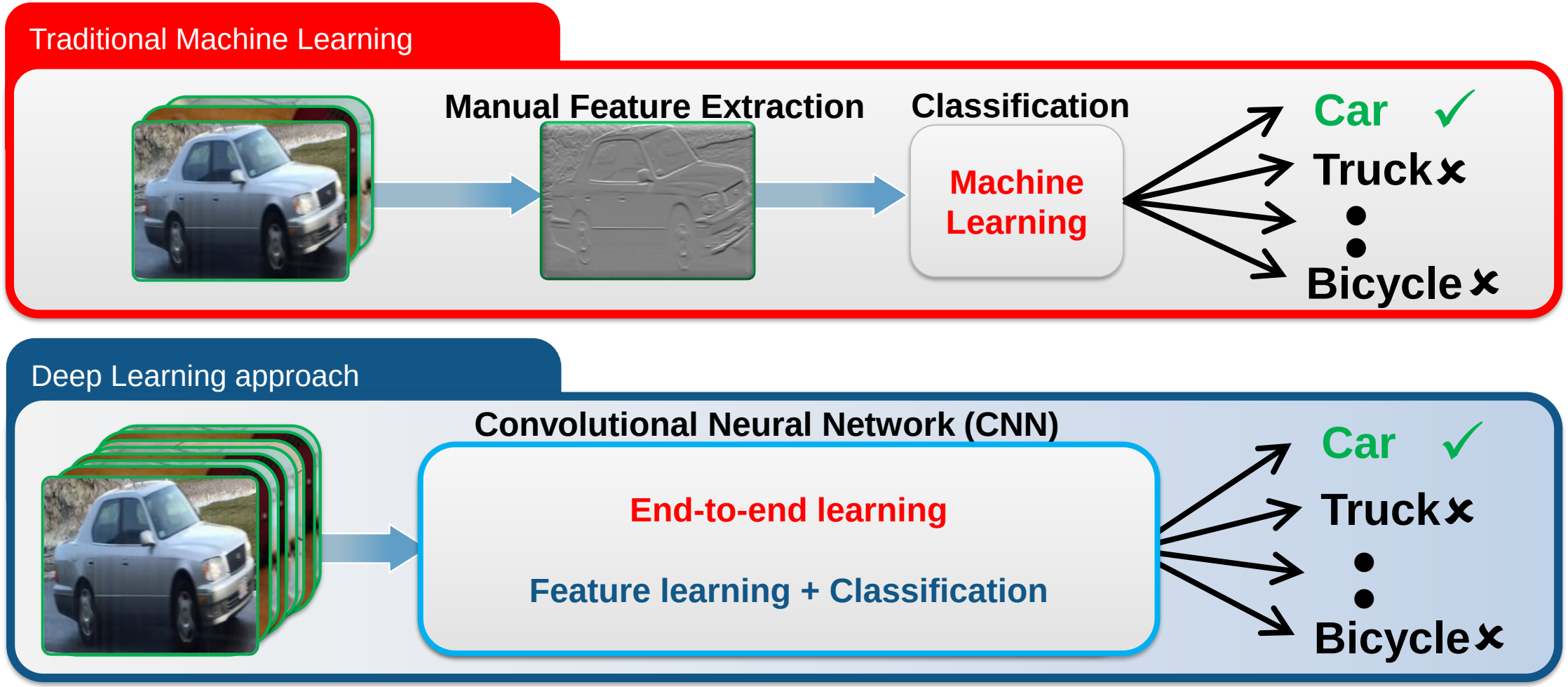


Robotics & Controls

and many more…

What is Deep Learning ?

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



Why is Deep Learning so Popular ?

- **Results:** Achieved substantially better results on ImageNet large scale recognition challenge
  - 95% + accuracy on ImageNet 1000 class challenge

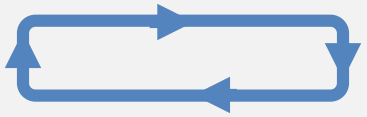
Year	Error Rate
Pre-2012 (traditional computer vision and machine learning techniques)	> 25%
2012 (Deep Learning)	~ 15%
2015 (Deep Learning)	<5 %

- **Computing Power:** GPU's and advances to processor technologies have enabled us to train networks on massive sets of data.

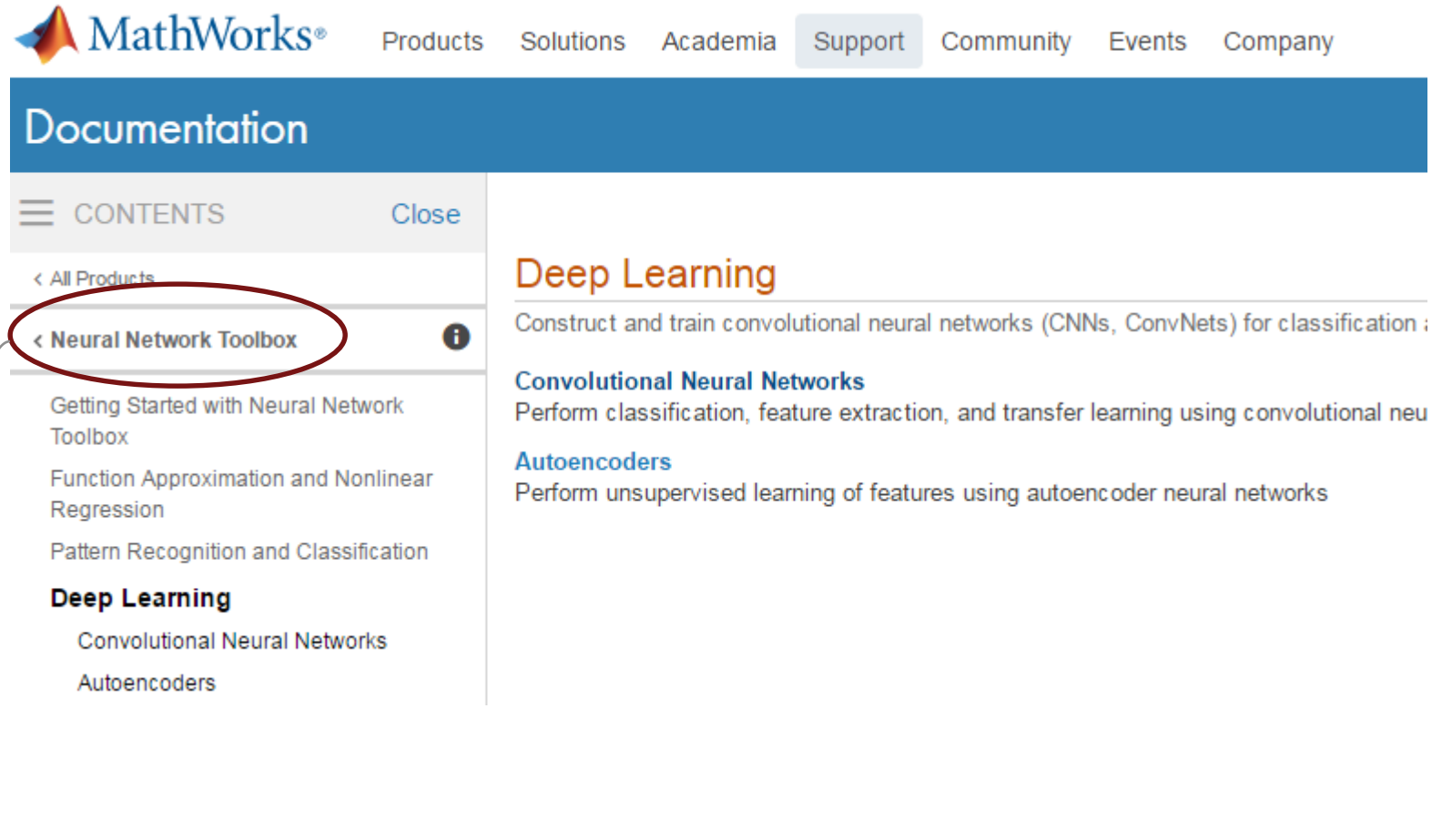
- **Data:** Availability of storage and access to large sets of labeled data
  - E.g. ImageNet , PASCAL VoC , Kaggle

How to build a Deep learning network?

Challenges using Deep Learning for Computer Vision

Steps	Challenge
Importing Data	Managing large sets of labeled images
Preprocessing	Resizing, Data augmentation
Choosing an architecture	Background in neural networks (deep learning)
Training and Classification	Computation intensive task (requires GPU)
Iterative design	

MATLAB does Deep Learning ?

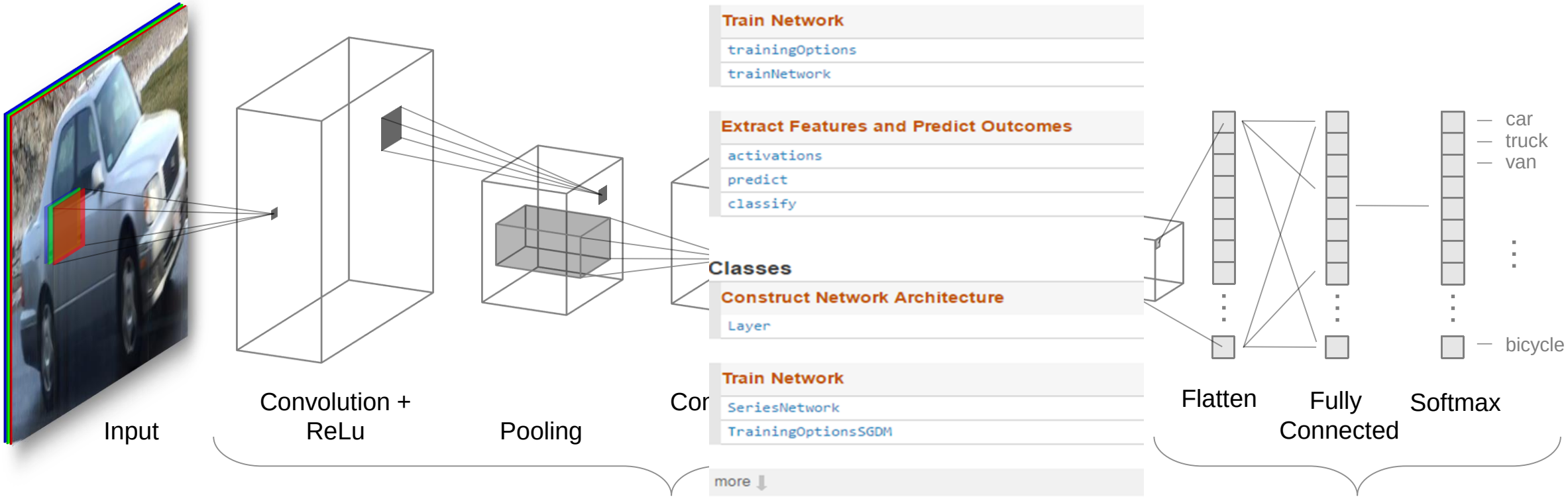


Neural Network Toolbox

Deep Learning

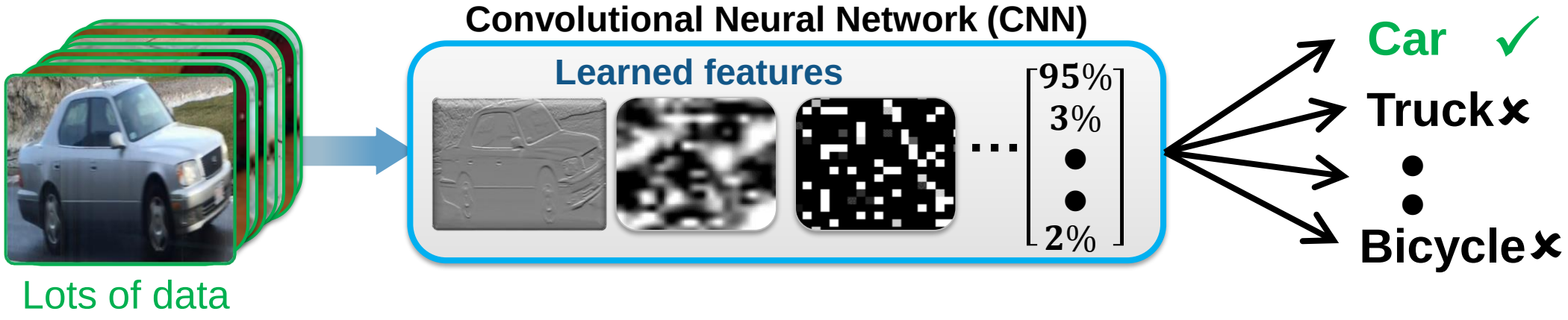
Convolutional Neural Networ

- Train “deep” neural networks on s ges, signals, text)
- Implements Feature Learning: Eli crafted” features
- Trained using GPUs for performar



Two Approaches for Deep Learning

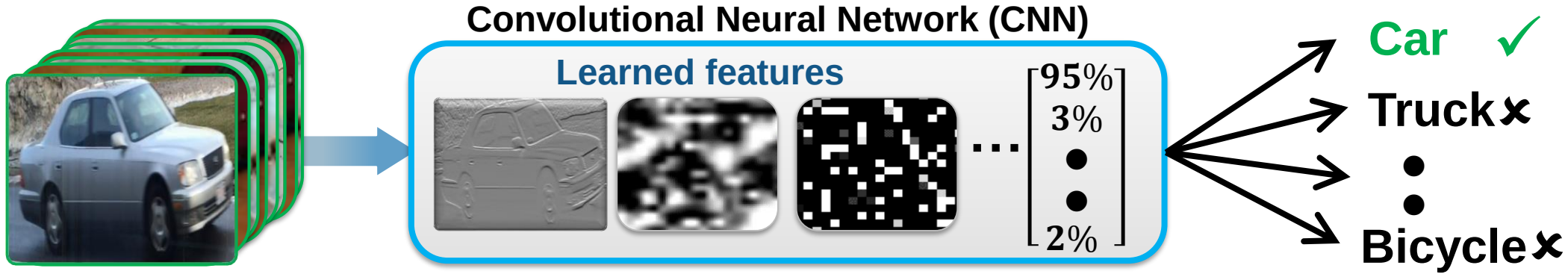
1. Train a Deep Neural Network from Scratch



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

Two Deep Learning Approaches

Approach 1: Train a Deep Neural Network from Scratch



Recommended <u>only</u> when:

Training data	1000s to millions of labeled images
Computation	Compute intensive (requires GPU)
Training Time	Days to Weeks for real problems
Model accuracy	High (can over fit to small datasets)

Demo: Classifying the CIFAR-10 dataset

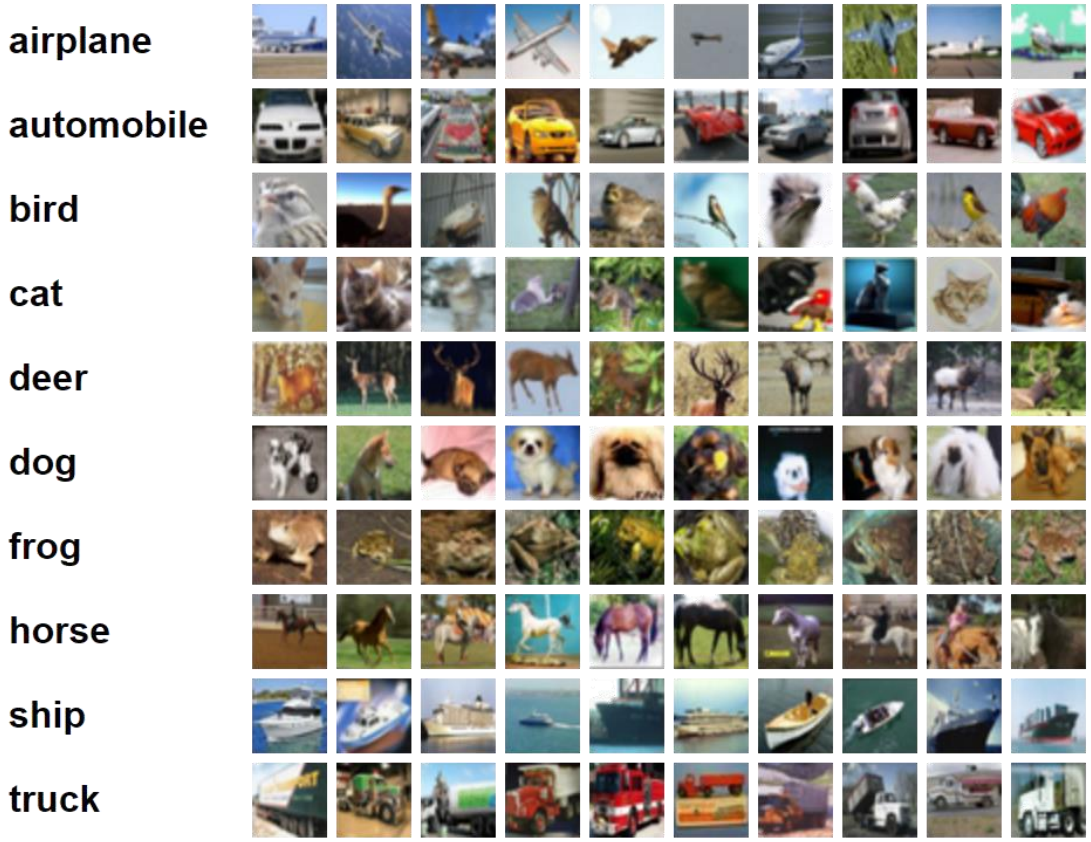
Objective: Train a Convolutional Neural Network to classify the CIFAR-10 dataset

Data:

Input Data	Thousands of images of 10 different Classes
Response	**AIRPLANE, AUTOMOBILE, BIRD, CAT, DEER, DOG, FROG, HORSE, SHIP, TRUCK**

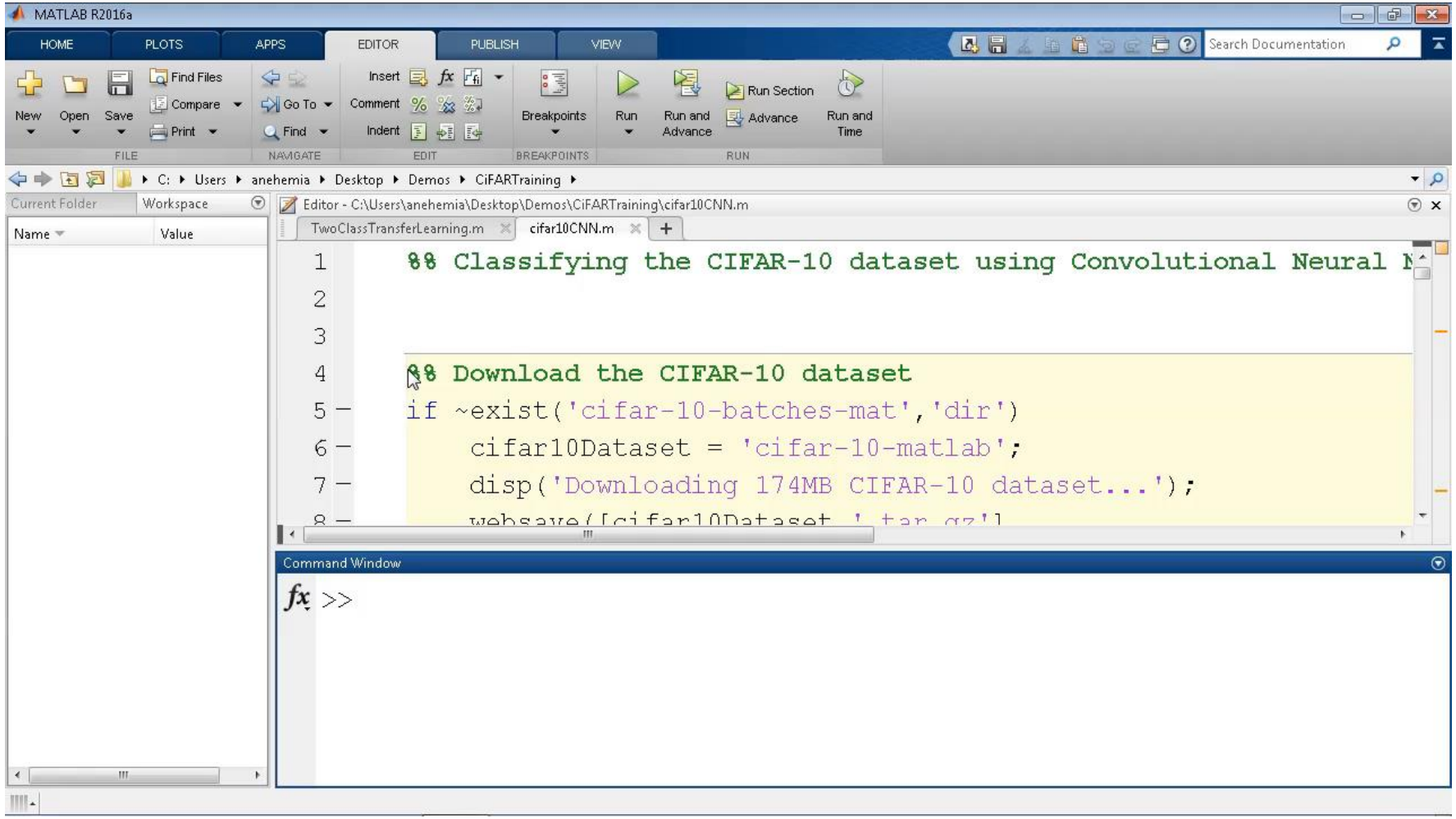
Approach:

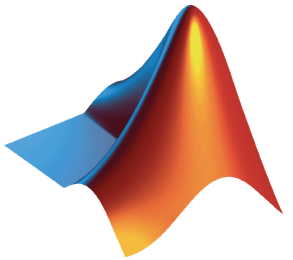
- Import the data
- Define an architecture
- Train and test the CNN



Data Credit: Learning Multiple Layers of Features from Tiny Images, Alex Krizhevsky, 2009. **https://www.cs.toronto.edu/~kriz/cifar.html**

Demo: Classifying the CIFAR-10 dataset





Two Deep Learning Approaches

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

CNN trained on massive sets of data

- Learned robust representations of images from larger data set
- Can be fine-tuned for use with ***new data or task*** with small – medium size datasets

Recommended when:

Training data	100s to 1000s of labeled images (small)
Computation	Moderate computation (GPU optional)
Training Time	Seconds to minutes
Model accuracy	Good, depends on the pre-trained CNN model

Demo

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

New Data

Pre-trained CNN
(AlexNet – 1000 Classes)

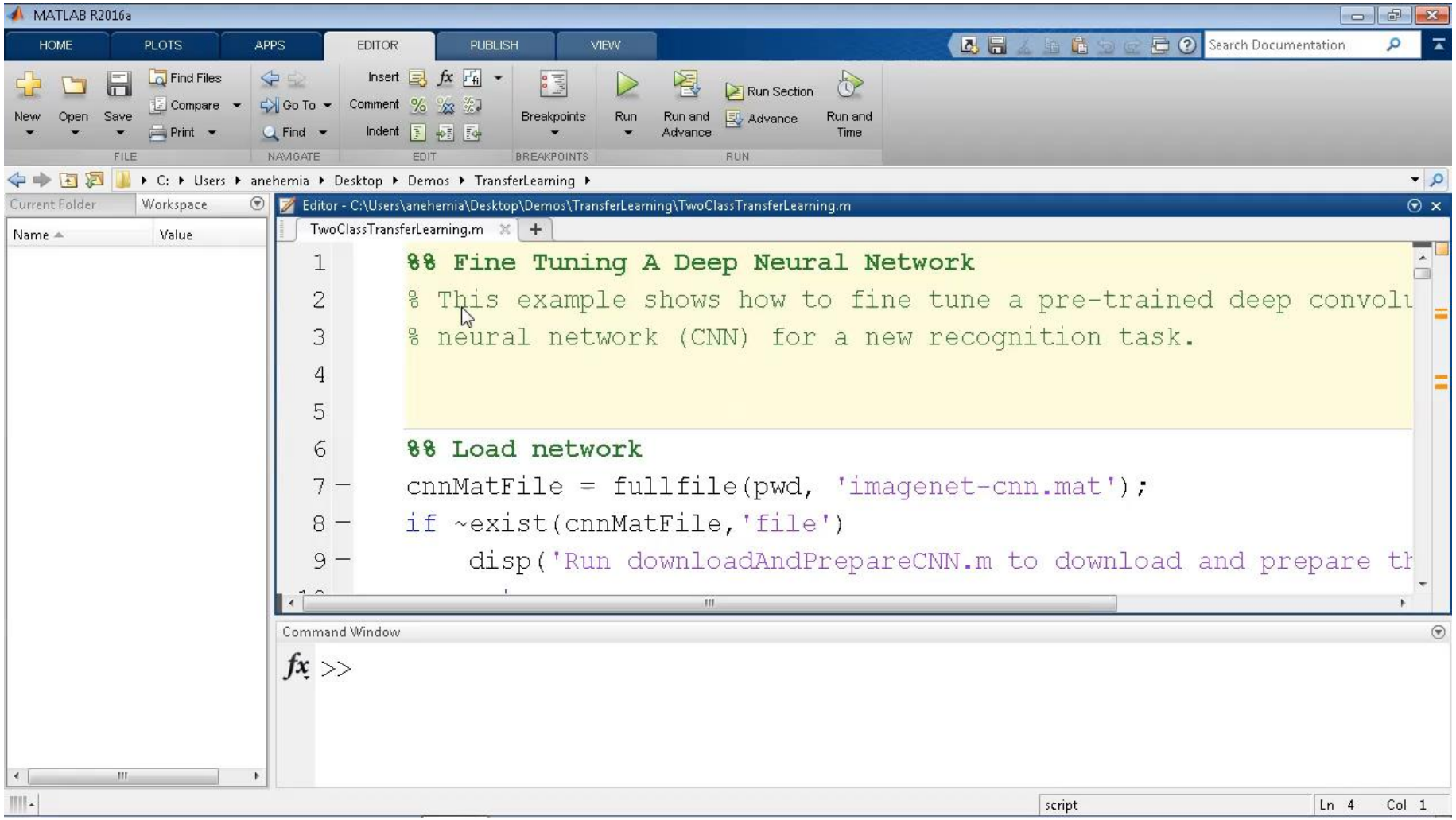
Car

SUV

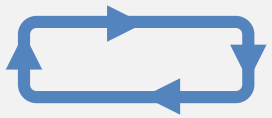
New Task – 2 Class
Classification

Demo

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



Addressing Challenges in Deep Learning for Computer Vision

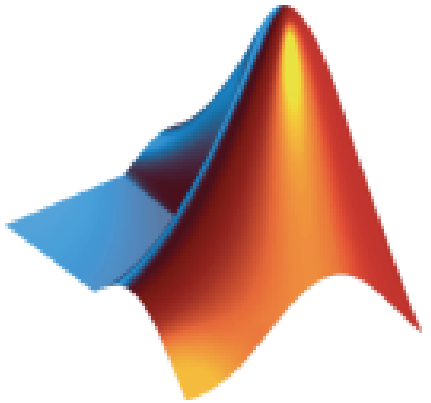
Challenge	Solution
Managing large sets of labeled images	**`imageSet`** or **`imageDataStore`** to handle large sets of images
Resizing, Data augmentation	**`imresize`**, **`imcrop`**, **`imadjust`**, **`imageInputLayer`**, etc.
Background in neural networks (deep learning)	Intuitive interfaces, well-documented architectures and examples
Computation intensive task (requires GPU)	Training supported on GPUs No GPU expertise is required
	Automate. Offload computations to a cluster and test multiple architectures

Key Takeaways

- Range of functionalities to support feature detection, 3D vision and camera calibration.

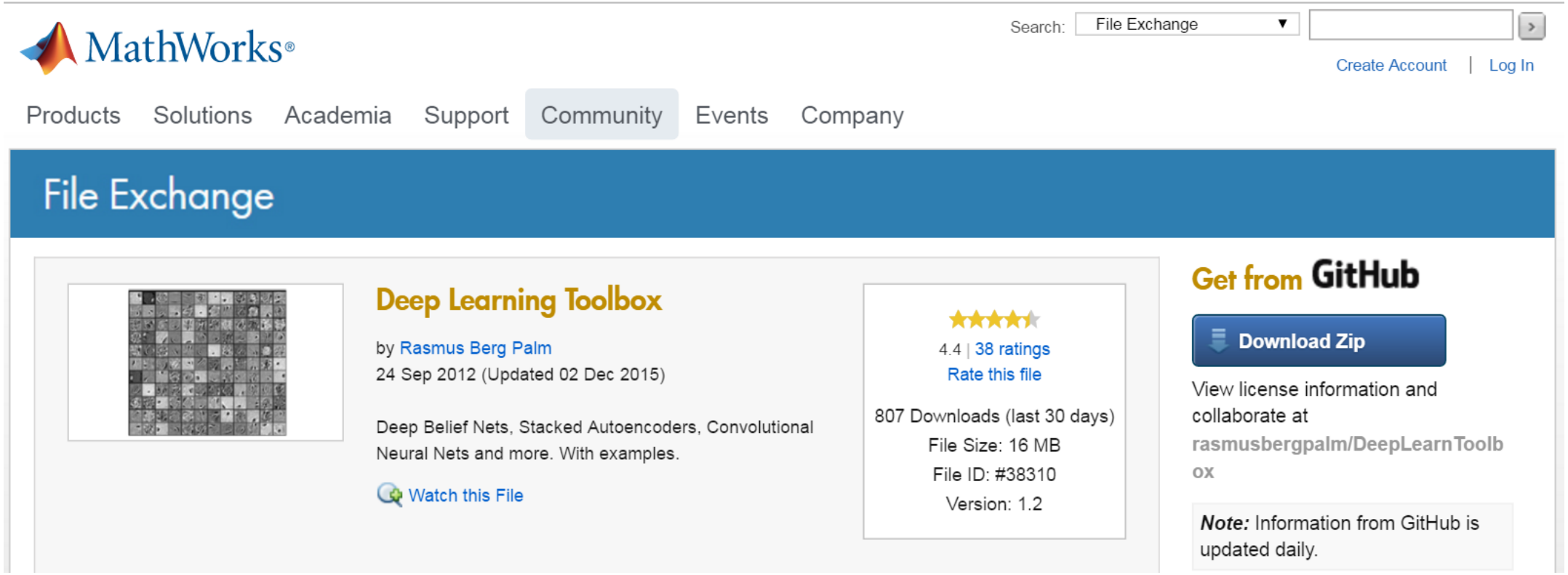
- Consider Deep Learning when:
  - Accuracy of traditional classifiers is not sufficient
    - *ImageNet classification problem*
  - You have a pre-trained network that can be fine-tuned
  - Too many image categories (100s – 1000s or more)
    - *Face recognition*

- Explore deep learning by building your own architecture in MATLAB



Further Resources on our File Exchange

- http://www.mathworks.com/matlabcentral/fileexchange/38310-deep-learning-toolbox



Thank You!

Questions?