

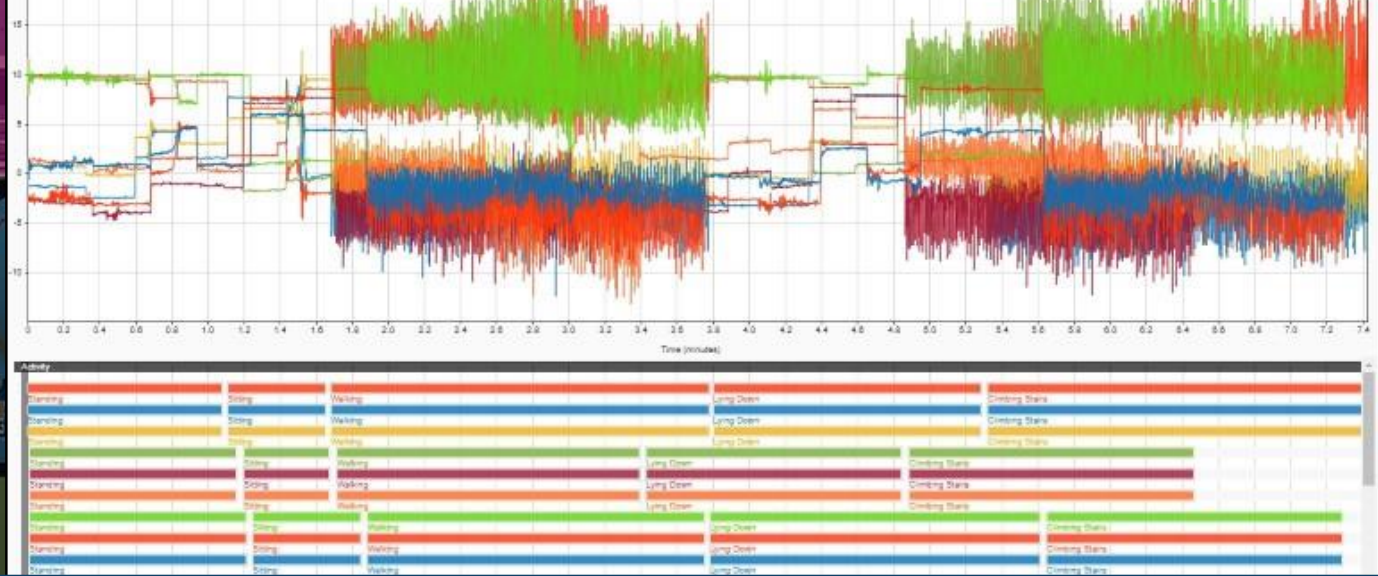
MATLAB과 함께하는 딥러닝 4주 완성 부트캠프

송완빈 과장

Application Engineer @ MathWorks

wsong@mathworks.com





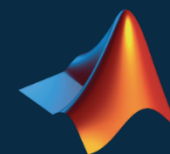
세션3. 딥러닝 프로젝트를 위한 데이터 준비 기법

MATLAB과 함께하는 딥러닝 4주 완성 부트캠프

송완빈 과장

Application Engineer @ MathWorks

wsong@mathworks.com



MathWorks®

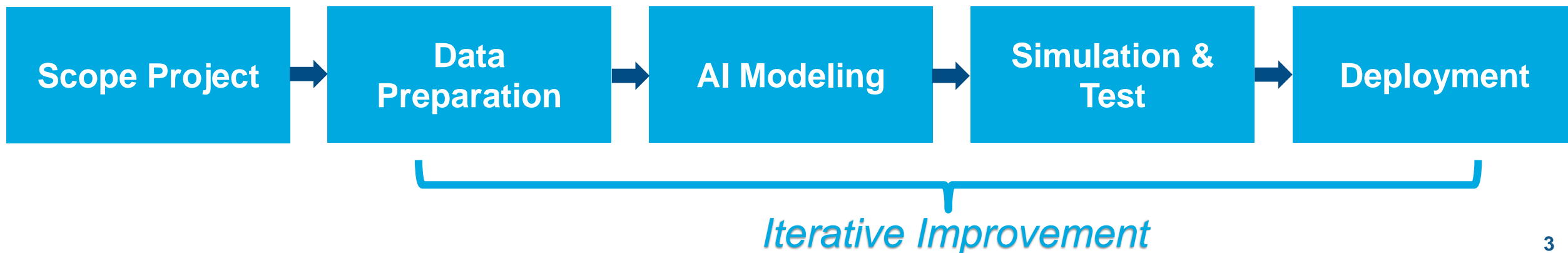
Accelerating the pace of engineering and science

Software System Design

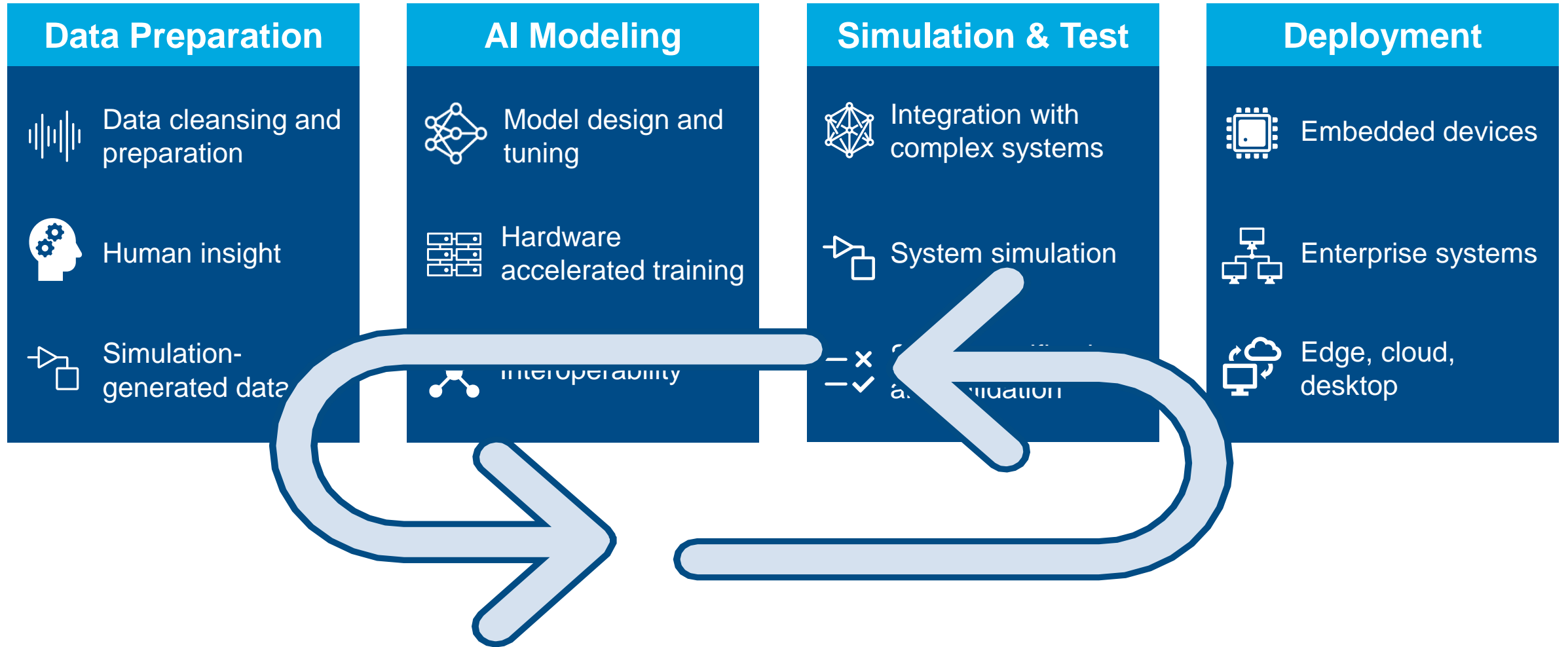
- Traditional Software



- AI Software = Code + Data
(Model/Algorithm)

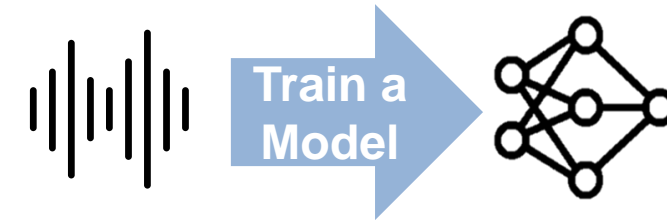
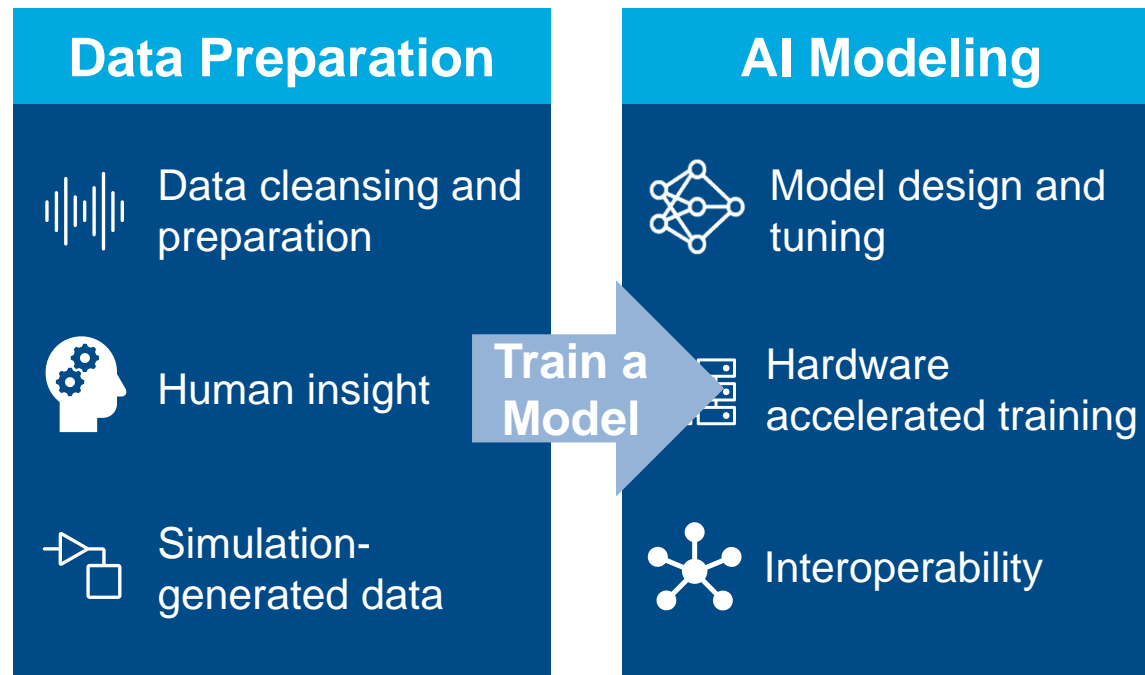


AI-driven system design



Iterative refinement for system improvement

Making it systematic : MLOps



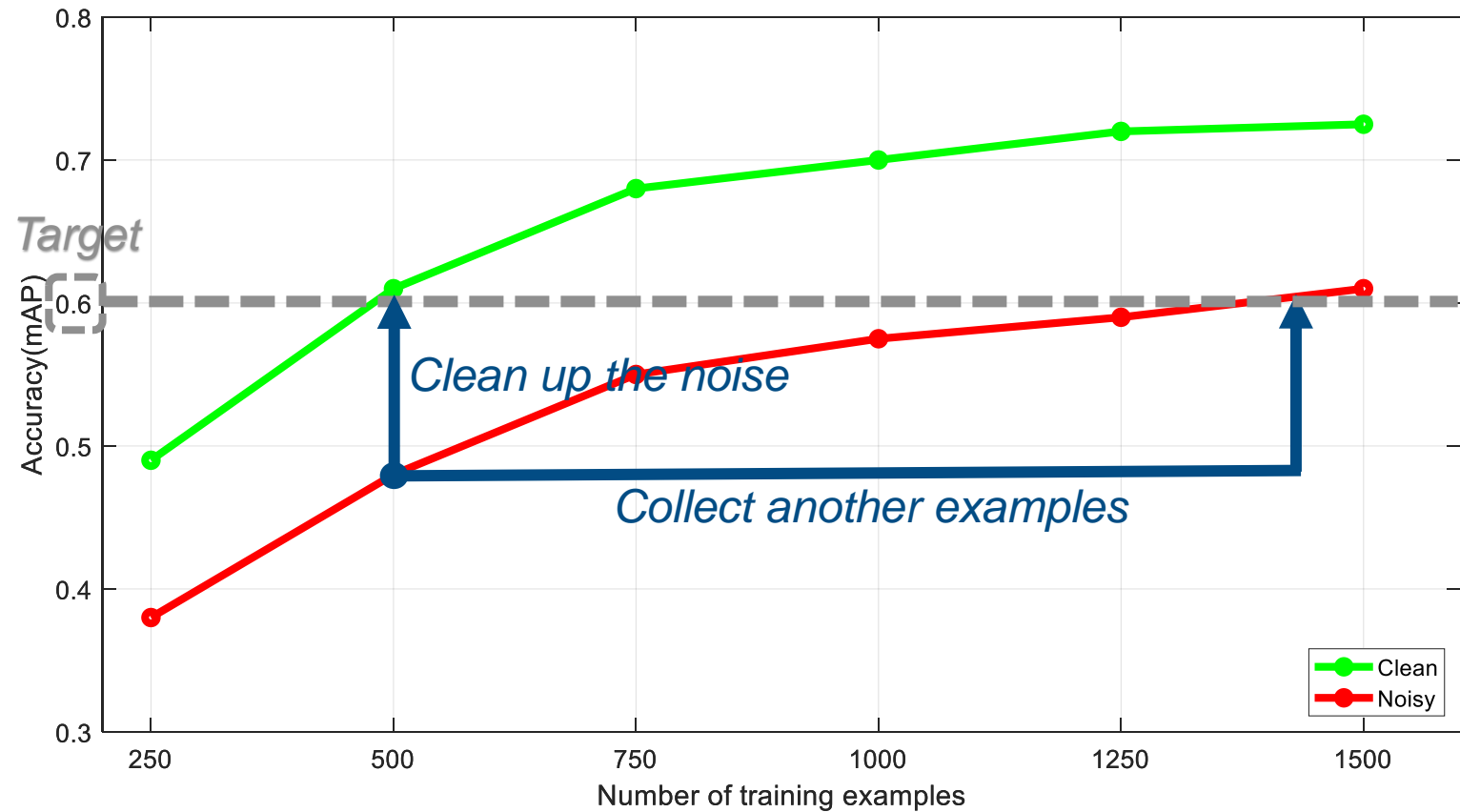
- Error Analysis
 - **Model-centric**
 - How can I tune the model architecture?
 - **Data-centric**
 - How can I modify my data?

- Data-centric AI
 - Error analysis to identify the types of data the algorithm does poorly on
 - Hold the code fixed and iteratively improve the data.

Data preparation represents most of your AI effort...

Transforming raw data for useful modeling and analysis is a critical step.

Example : Clean vs. Noisy data



Source: Andrew Ng slide from MLOps: From Model-centric to Data-centric AI 2021

Data Preparation

- Data cleansing and preparation
- Human insight
- Simulation-generated data

Data preparation represents most of your AI effort

Transforming raw data for useful modeling and analysis is a critical step.

Example : Clean vs. Noisy data

Data Preparation

- Data cleansing and preparation
- Human insight
- Simulation-generated data



Big Data to Good Data

From Big data to Good Data

Transforming raw data for useful modeling and analysis is a critical step.

Data

(inputs x , labels y) ***Big Data to Good Data***

Data Preparation



Data cleansing and preparation



Human insight



Simulation-generated data

- Good data is:
 - Defined consistently
 - Definition of labels y is unambiguous
 - Cover of important cases
 - Good coverage of inputs x
 - Has timely feedback from production data
 - Distribution covers data drift and concept drift
 - Sized appropriately

Synthetic Data Generation and Augmentation to deal with less data

Simulate your system for synthetic data generation

Data
(**inputs x**, labels y)

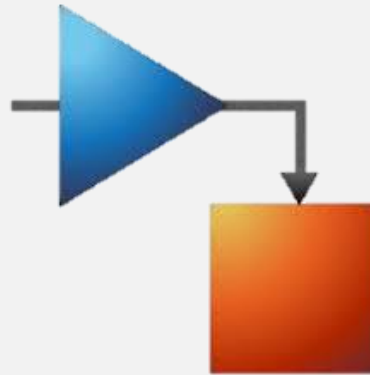
Data Preparation

Data cleansing and preparation

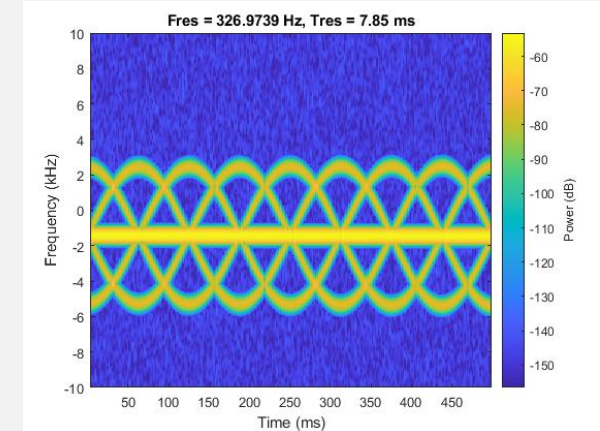
Human insight

Simulation-generated data

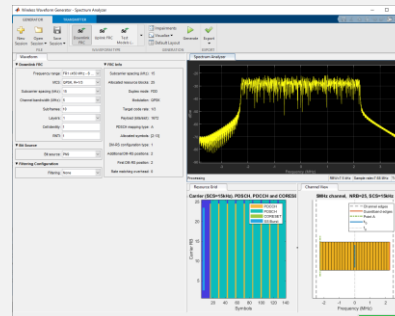
Simulate data using models and deep learning



Generate Radar Returns

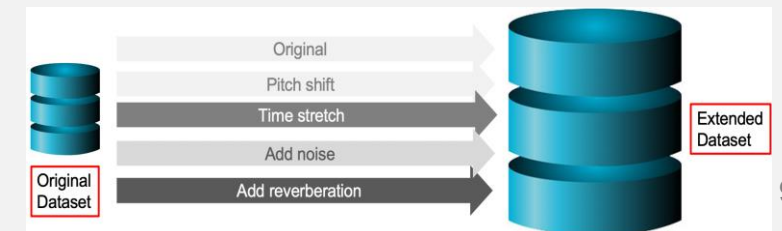


Generate wireless waveforms



Generate Audio Data

text2speech




Synthetic Data Generation and Augmentation to deal with less data

Simulate your system for synthetic data generation


Data

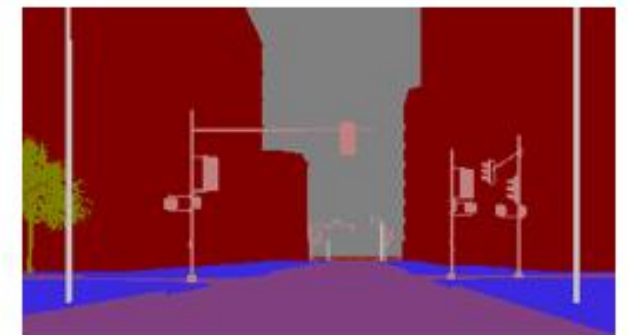
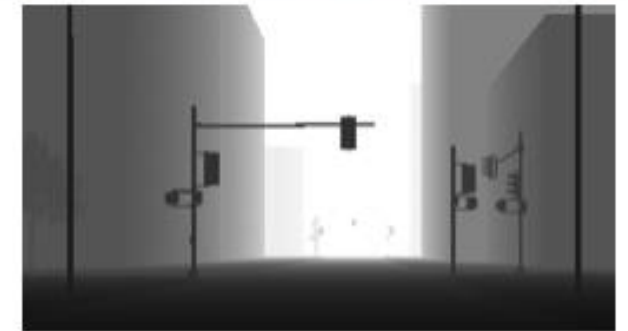
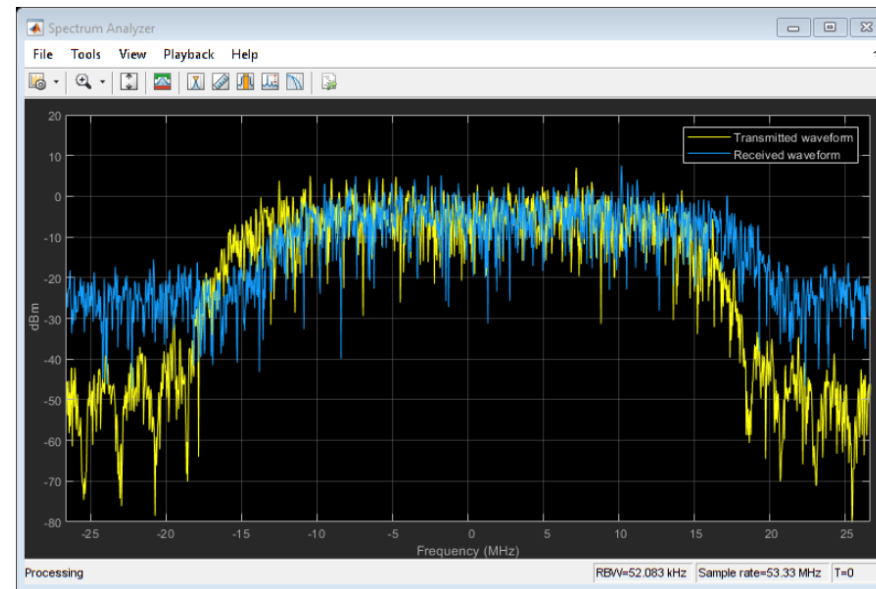
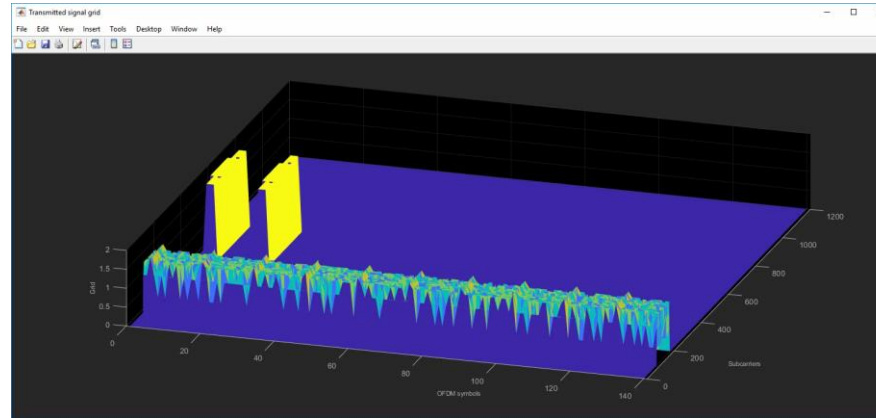
(inputs x , labels y)

Data Preparation

 Data cleansing and preparation

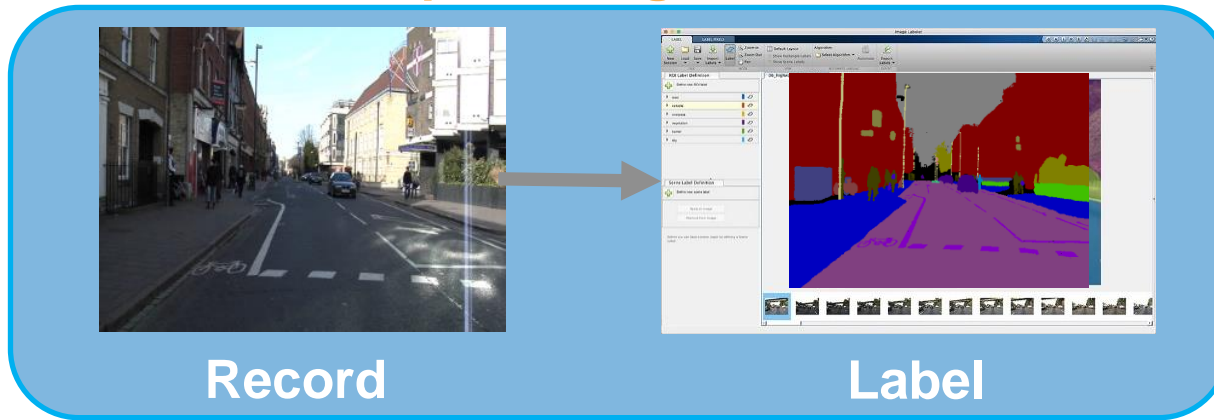
 Human insight

 Simulation-generated data

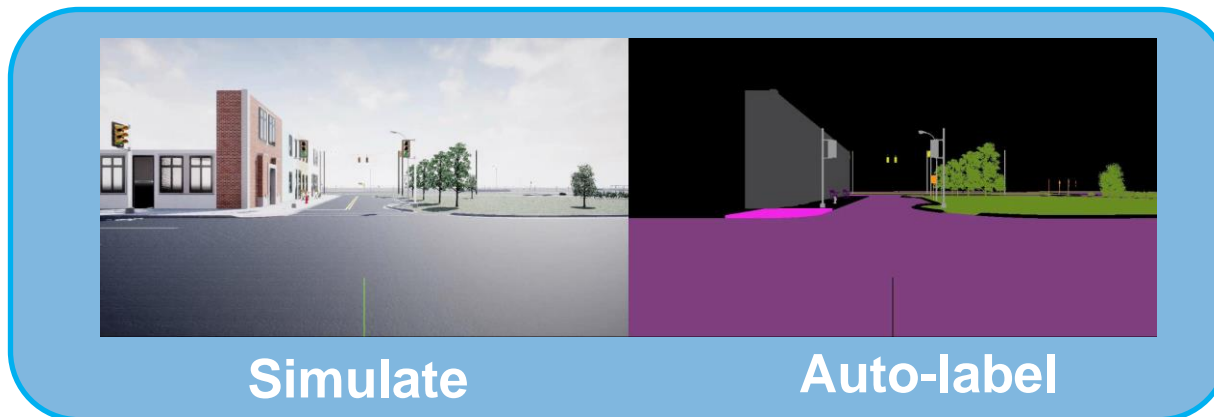


One example of leveraging simulation for data synthesis

Traditional deep learning workflow



Simulation-based workflow



Spend less time preprocessing and labeling data

Synchronize disparate time series, filter noisy signals.

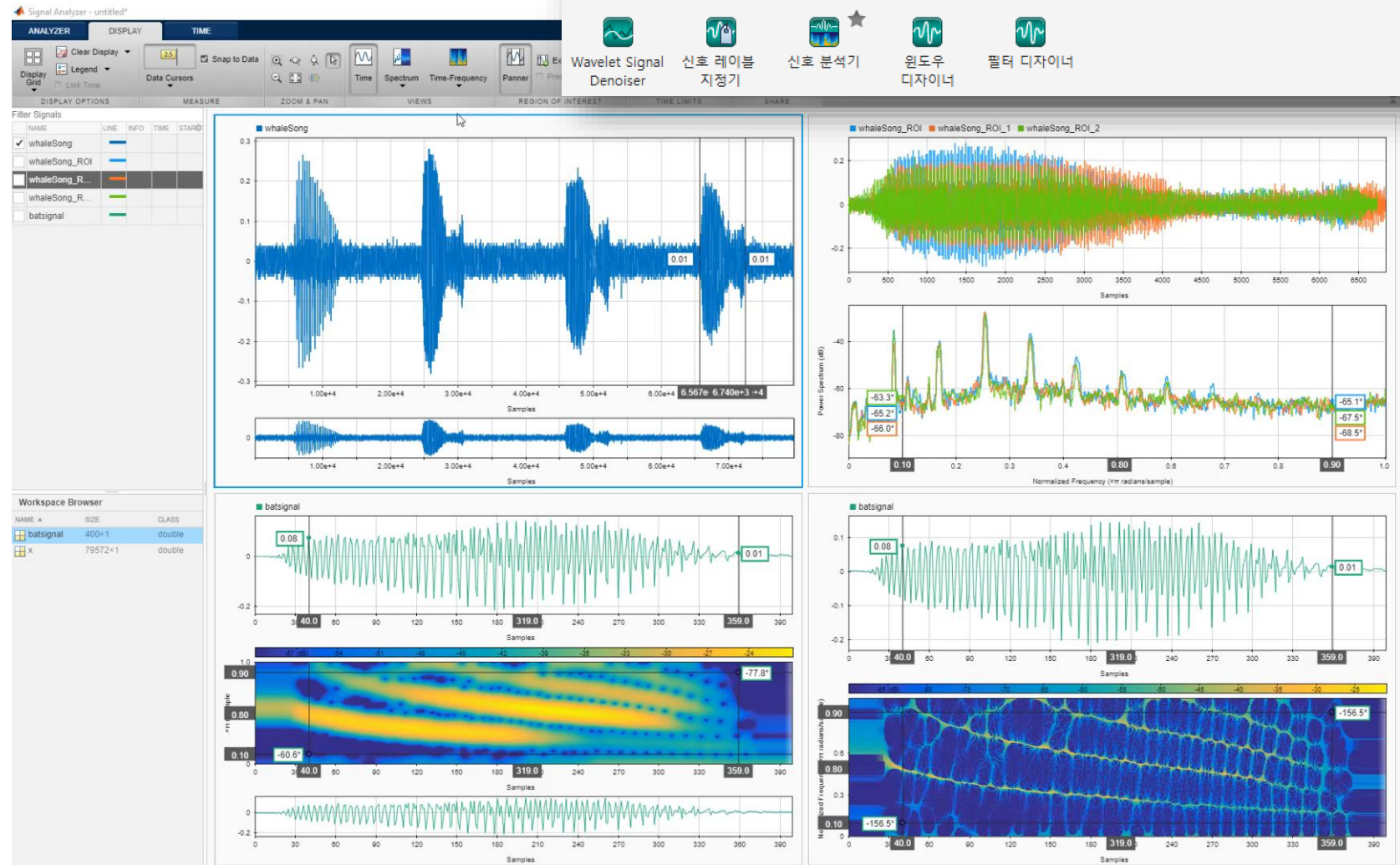
Data

(inputs x , labels y)

신호 처리 및 통신

Toolbox icons for Signal Analyzer:

- Audio Labeler
- Audio Test Bench
- EDF 파일 분석기
- Filter Builder
- Impulse Response M...
- Pulse Waveform A...
- Radar Designer
- Sensor Array Analyzer
- Signal Multiresoluti...
- Wavelet Signal Denoiser
- 신호 레이블 지정기
- 신호 분석기
- 윈도우 디자이너
- 필터 디자이너



Data Preparation

Data cleansing and preparation

Human insight

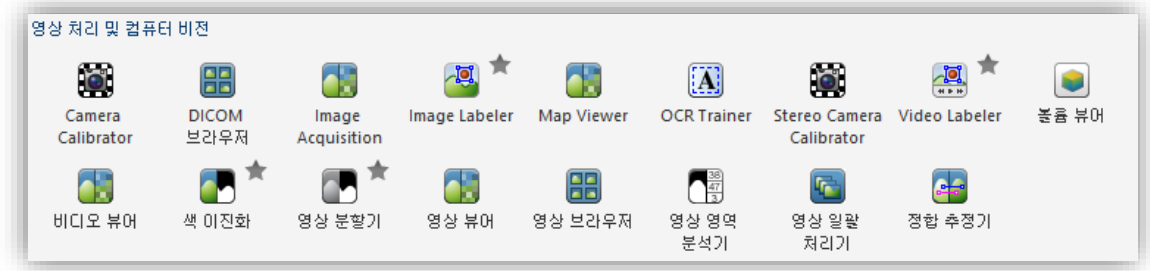
Simulation-generated data

Spend less time preprocessing and labeling data

Preprocess image data using built-in Apps

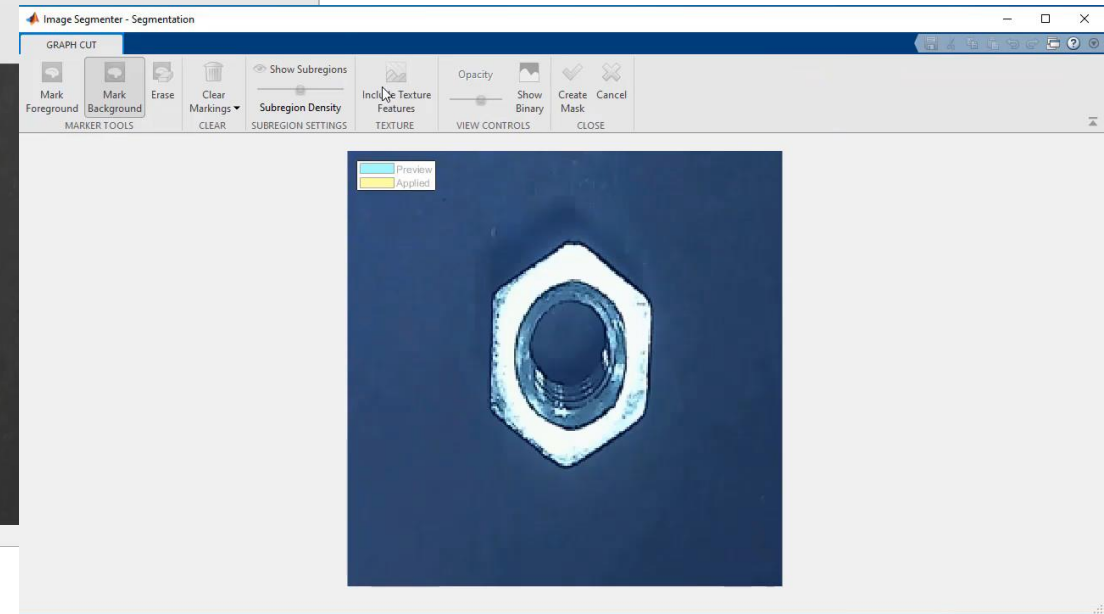
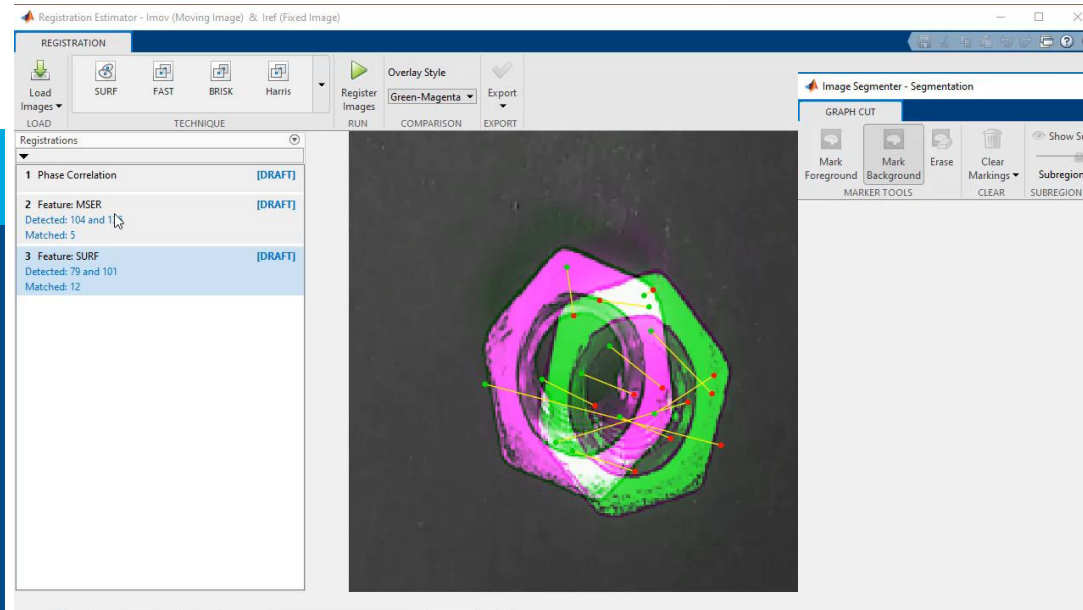
Data

(inputs **x**, labels **y**)

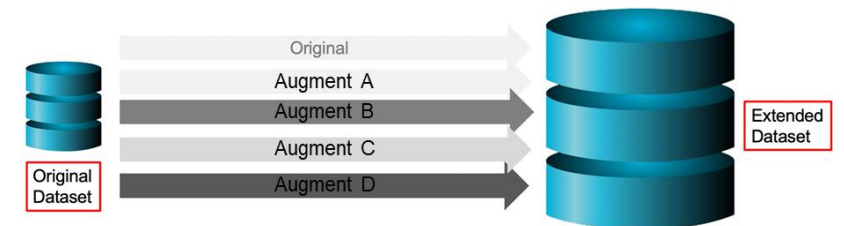


Data Preparation

- Data cleansing and preparation
- Human insight
- Simulation-generated data



```
>> audioDataAugmenter
>> imageDataAugmenter
```



Spend less time preprocessing and labeling data

Automate labeling of Lidar, image, video, and signal.

Data
(inputs x , labels y)

Data Preparation



Data cleansing and preparation



Human insight

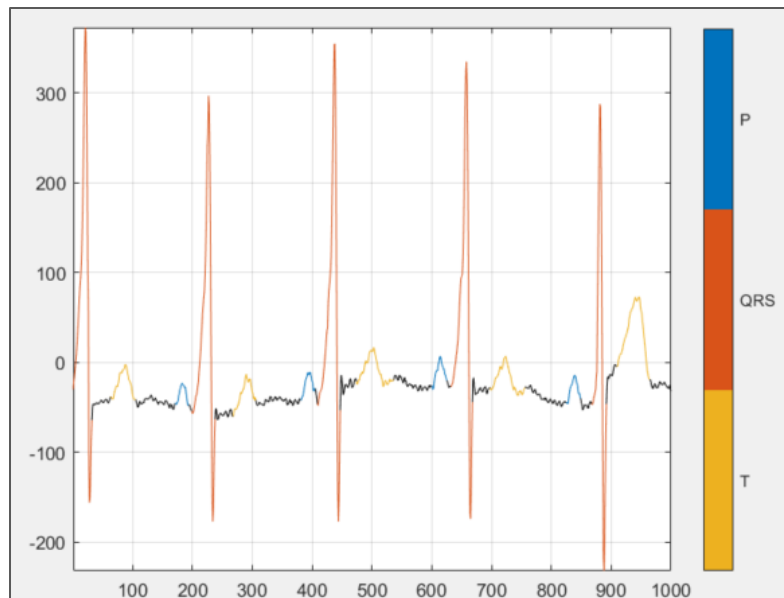
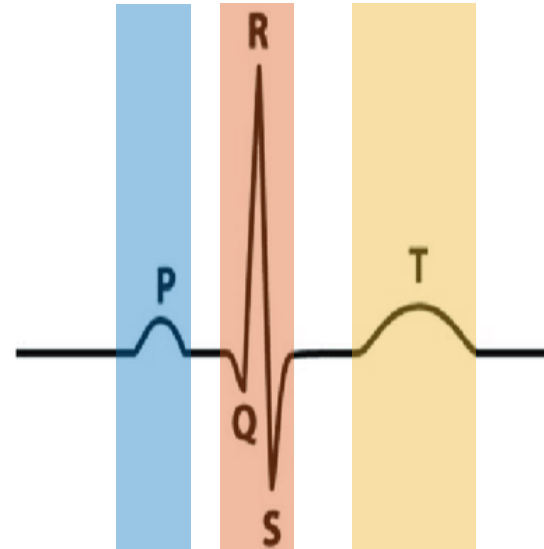
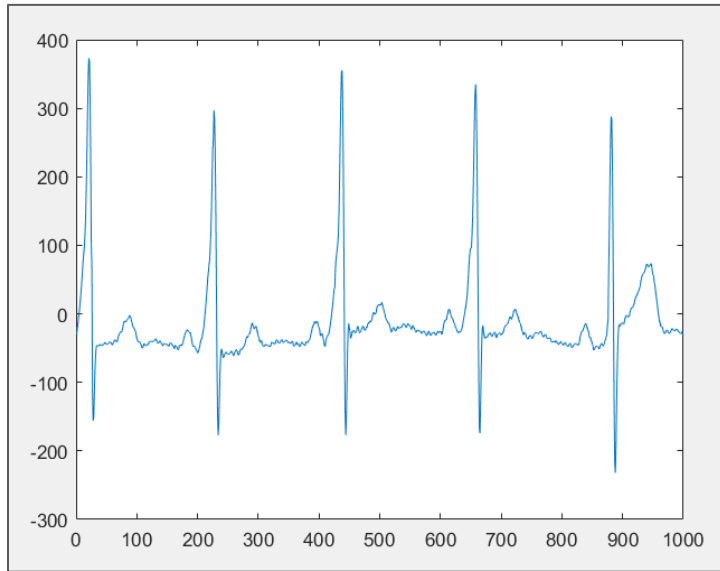


Simulation-generated data

The screenshot displays the Lidar Labeler software interface. The main window is titled 'Lidar Labeler' and features a toolbar with options like 'Shrink To Fit', 'Snap to Cluster', 'Colormap', and 'Ego Direction'. Below the toolbar are tabs for 'GROUND', 'CUBOID', 'COLORMAP', and 'CAMERA VIEW'. The central area shows a 3D point cloud of a scene with a car labeled as 'vehicle'. To the left is a 'ROI Labels' panel with a tree view showing 'vehicle' (containing 'car', 'bike', 'pole', 'vegeta...') and 'None'. Below this is a 'Scene Label Definition' section. In the foreground, an 'Image Labeler' window shows a 2D image of a car with a yellow bounding box. To the right, a 'Labeled Signal Set Browser' table is visible, listing signal names, values, and locations. Below the table are three plots: a 'SPECTROGRAM' showing frequency vs. time, a 'Time' plot showing amplitude vs. time, and a 'Fast Fourier Transform' plot showing frequency vs. time.

Name	Plot	Value	Locatio...	Locatio...	Time
s	<input checked="" type="checkbox"/>				Fs: 14.00...
Fextrema					
<input checked="" type="checkbox"/>	1	1.054878...	1.054878...		
<input checked="" type="checkbox"/>	2	1.310975...	1.310975...		
<input checked="" type="checkbox"/>	3	1.524390...	1.524390...		
<input checked="" type="checkbox"/>	4	1.719512...	1.719512...		
<input checked="" type="checkbox"/>	5	1.884146...	1.884146...		
<input checked="" type="checkbox"/>	6	0.728658...	0.728658...		
<input checked="" type="checkbox"/>	7	0.341463...	0.341463...		
Atom					
<input checked="" type="checkbox"/>	1	0.149390...	0.259146...		
<input checked="" type="checkbox"/>	2	0.448170...	0.554878...		
<input checked="" type="checkbox"/>	3	0.945121...	1.060678...		
<input checked="" type="checkbox"/>	4	1.698170...	1.801829...		

Example1 : Labeling key regions of ECG waveforms



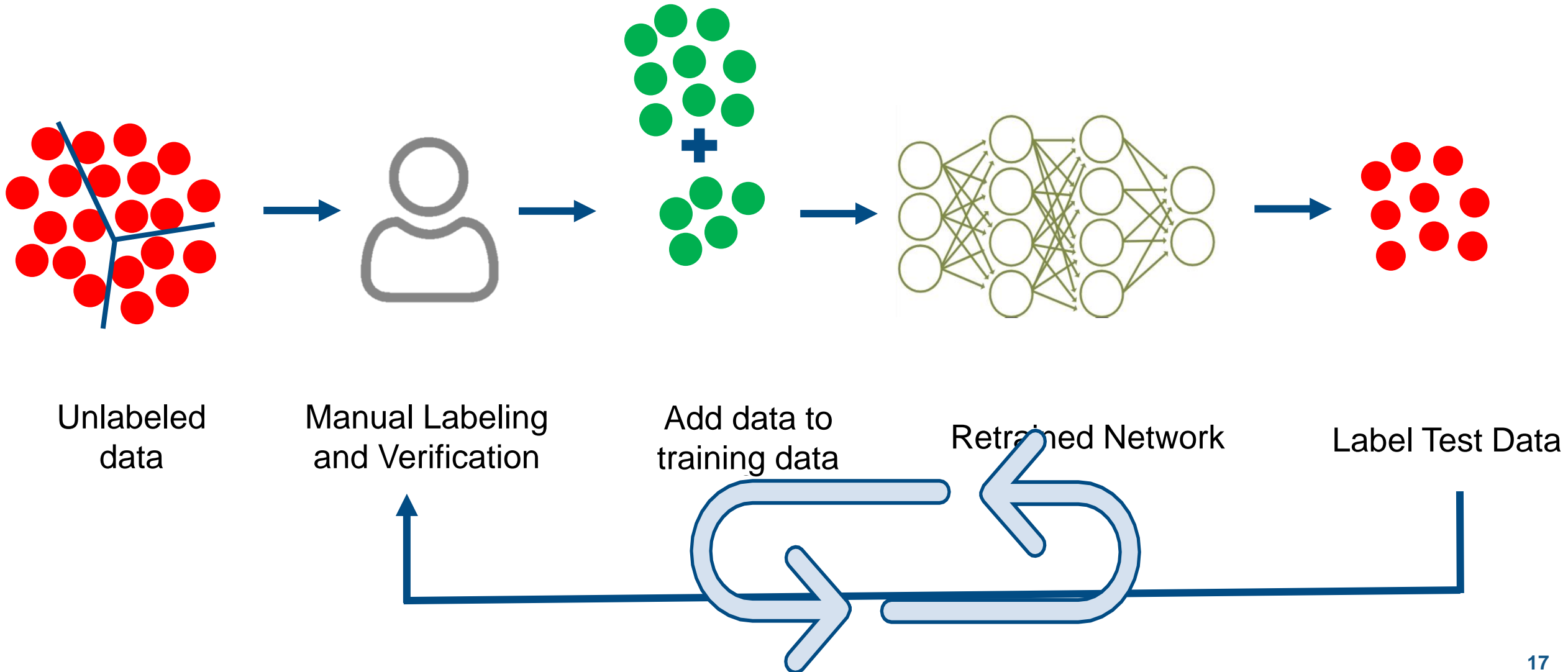
- ECG signals contain P, QRS and T waves.
- Identifying these can help with diagnosis and classification.
- Dataset contains 210 ECG signals, ~ 15 minutes long, labeled by cardiologist

Example2 : Labeling vehicles for object detection



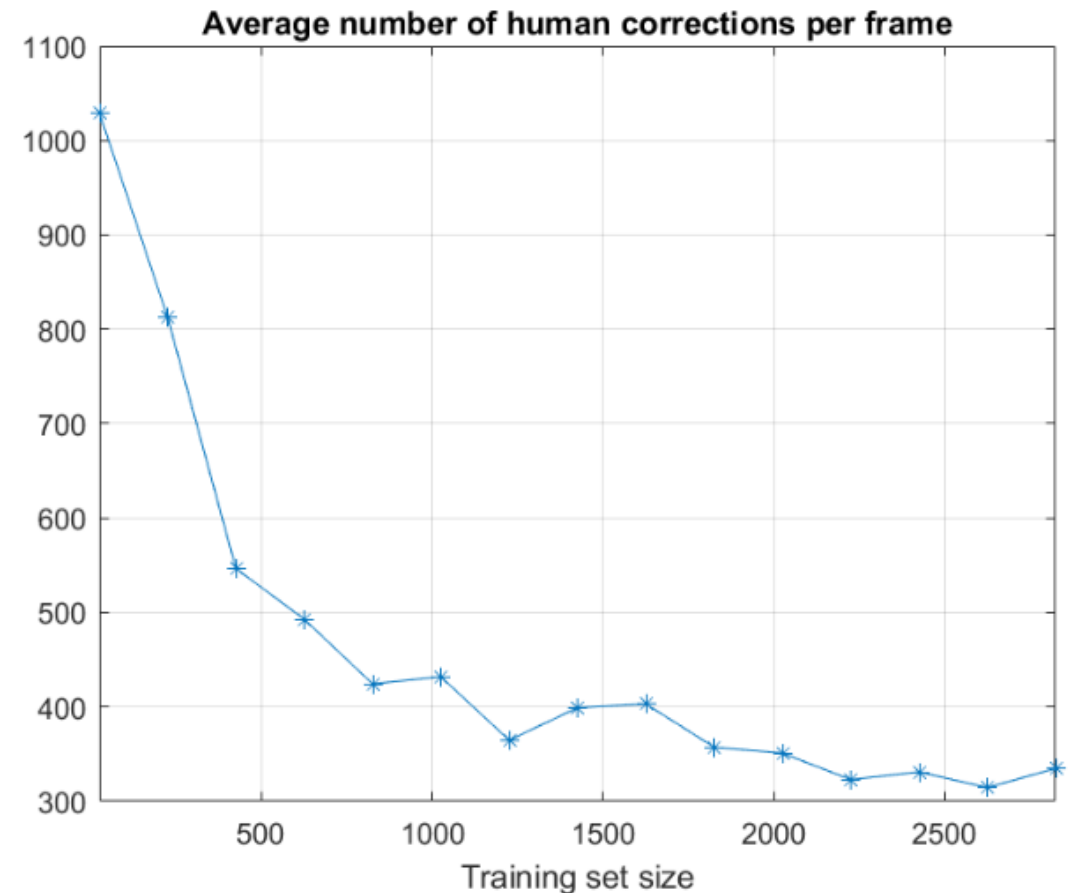
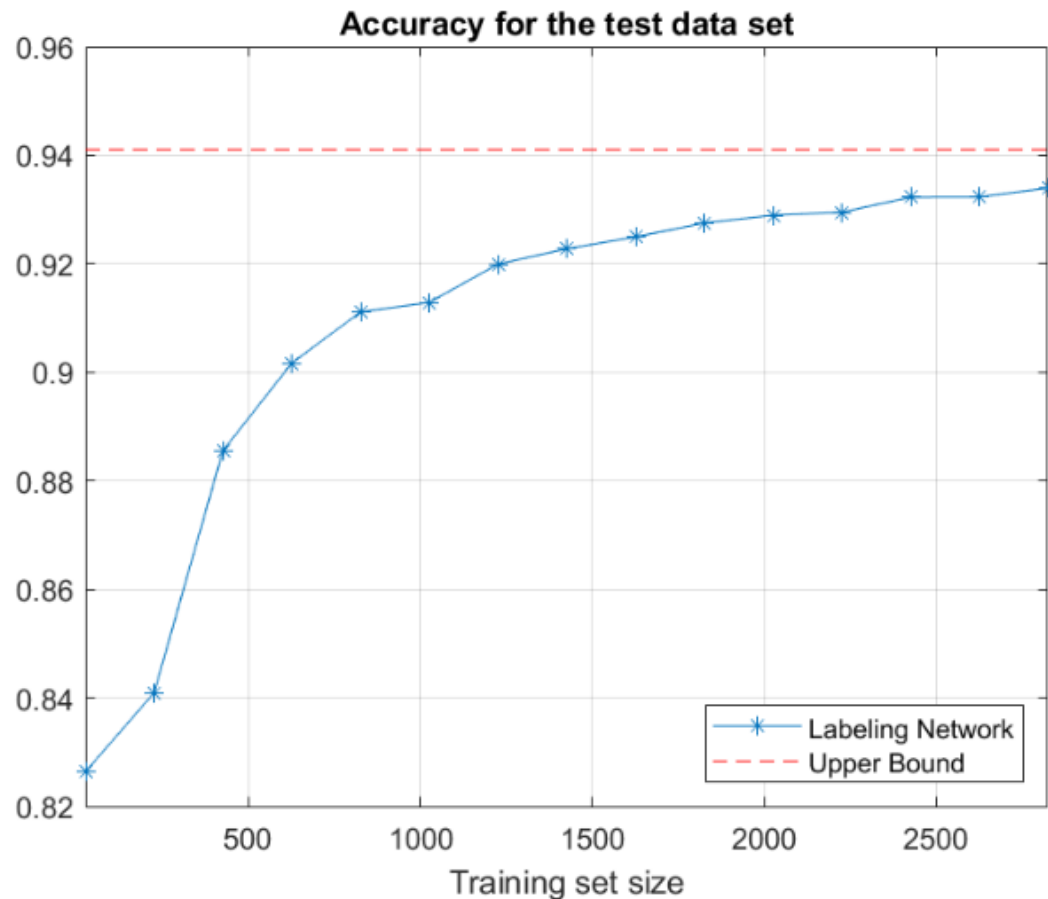
- Video data from my black box
- Wants to label cars in every frames of the video
- Do not have experiences on custom algorithm to detect vehicle
- Attended MATLAB Deep Learning Bootcamp #1

Establish iterative labeling process with deep learning



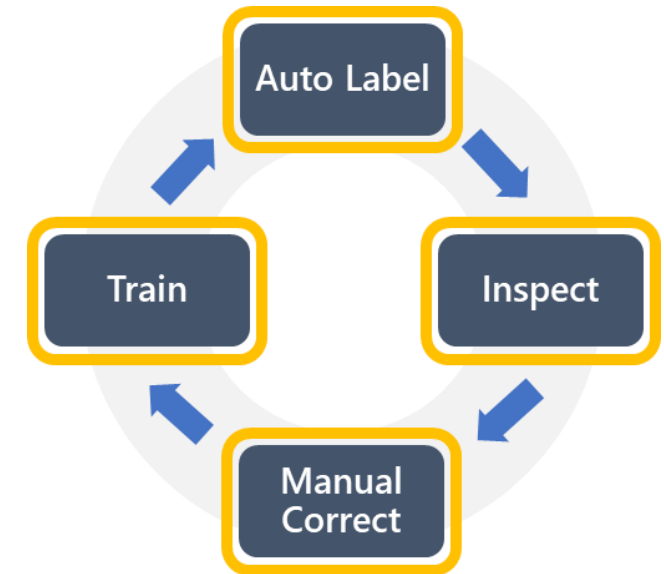
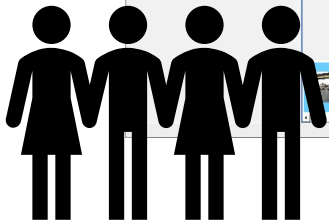
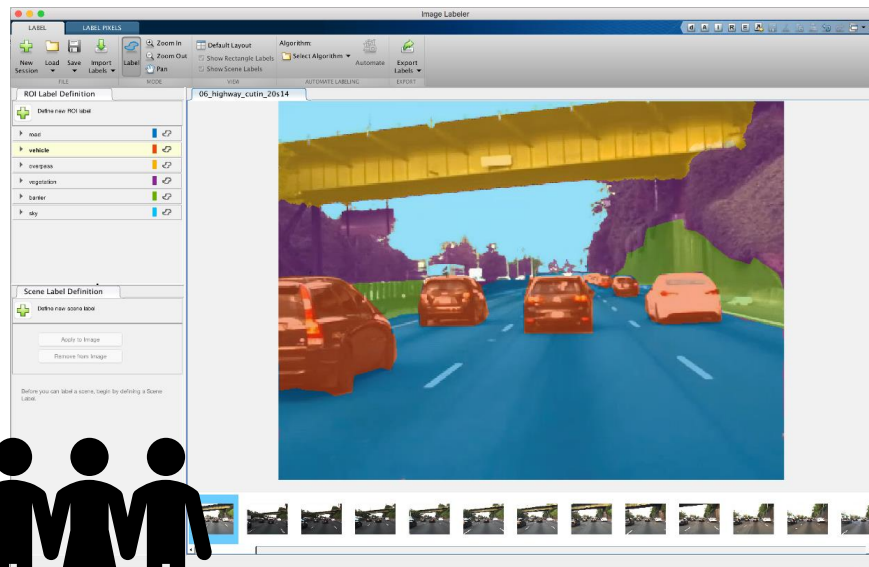
Result of iterative learning process

- Iteratively the accuracy increases, needs for human correction decreases.



Strategy for automated labeling and iterative learning

- Manually label with colleagues
- Use pretrained model for labeling automation
- Establish iterative labeling process



HYUNDAI STEEL uses pixel-based deep learning technique for steel material image



HYUNDAI STEEL developed pixel-based steel material image analysis application using deep learning and established labeling automation, enhancement technique with MATLAB.

In the field, analysis engineers using dot method, which manually analyze images from an optical microscope but it takes lots of time and strongly depends on the experience of the individual analyst.

With MATLAB, they created labeling automation algorithm using unsupervised machine learning technique and integrated it to labeling app for automation. By iterating labeling data based on machine learning and training network with deep learning model, they've enhanced prediction accuracy up to 85% which can fairly acceptable for replacing current analysis workflow.

Advantages of using MATLAB and Simulink:

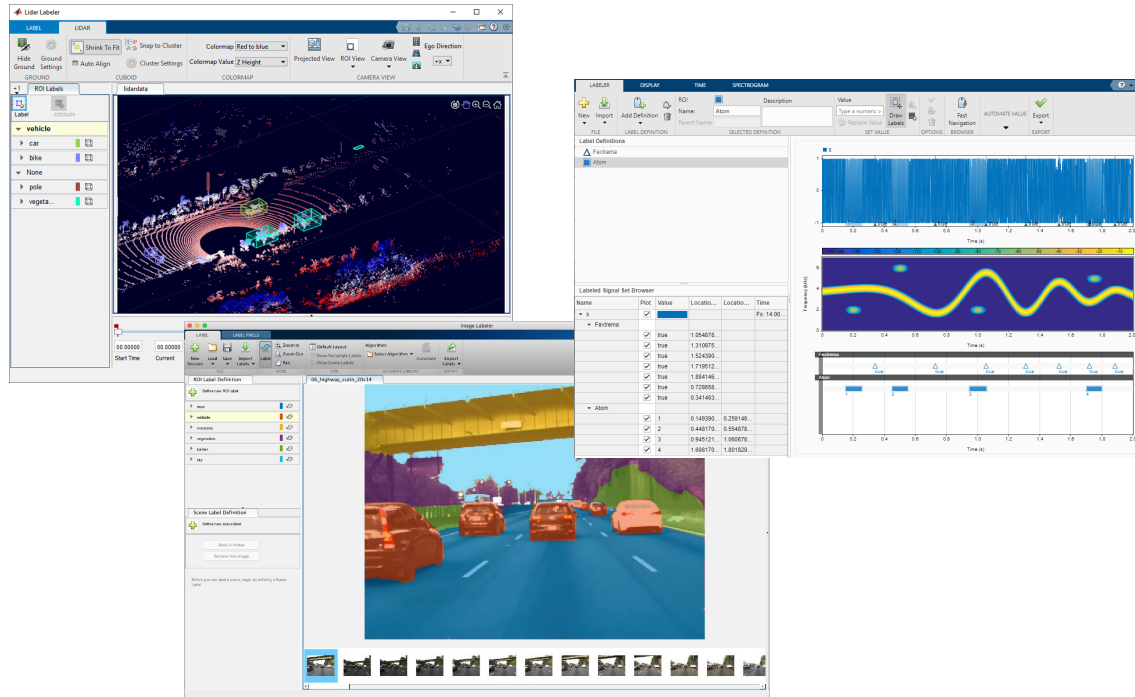
- Leveraged high-level APIs for applying image preprocessing techniques
- Interactive Apps enabled non-AI expert can do deep learning project
- Golden reference examples for quick prototyping
- Leverage MathWorks support to maximize the benefits of using MATLAB

“ Even though I had limited knowledge on Image processing and deep learning, I could successfully adopt deep learning for my project. With evaluation support from MathWorks, we could prototype our approach easily with limited time bound.

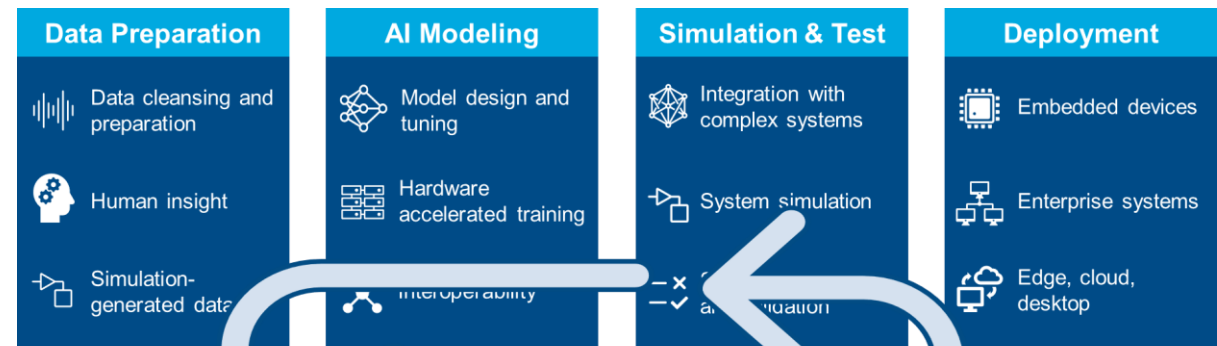
”



Build Systematic Data Management Process



	Predictive Maintenance • Bearing Prognosis • Pump Fault Diagnosis	Predictive Maintenance Toolbox™
	Land-Use Classification • Semantic Segmentation for Multispectral Images	Image Processing Toolbox™
	Lidar • Lidar Point Cloud Semantic Segmentation • 3-D Object Detection Using PointClouds	Lidar Toolbox™
	Radar • Radar Waveform Classification • Pedestrian and Bicyclist Classification	Phased Array System Toolbox™
	Wireless Communications • Modulation Classification • Detect WLAN Router Impersonation	Communications Toolbox™
	Reinforcement Learning • Train Biped Robot to Walk • PMSM Motor Control	Reinforcement Learning Toolbox™
	Computational Finance • Machine Learning for Statistical Arbitrage	Financial Toolbox™
	Robotics • Avoid Obstacles using Reinforcement Learning	Robotics System Toolbox™
	Automated Driving • Deep Learning Vehicle Detector • Occupancy Grid with Semantic Segmentation	Automated Driving Toolbox™
	Visual Inspection • Manufacturing Defect Detection • Anomaly Detection for Cloth Manufacturing	Image Processing Toolbox™
	Audio • Speech Command Recognition • Cocktail Party Source Separation	Audio Toolbox™
	Medical Imaging • 3-D Brain Tumor Segmentation • Breast Cancer Tumor Classification	Image Processing Toolbox™



- App based labeling tools
- Application specific preprocessing functions
- Deployment workflow for timely feedback from production data

MathWorks: helping engineers & scientists build Deep Learning solutions



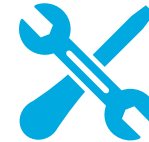
The Platform

MATLAB, Simulink, and over 100 add-on products for specialized applications



Your People

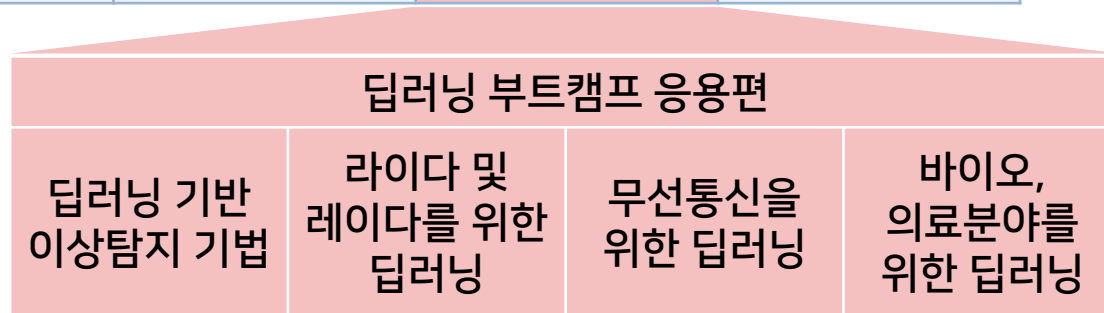
Helping you build an agile workforce today and preparing tomorrow's engineers



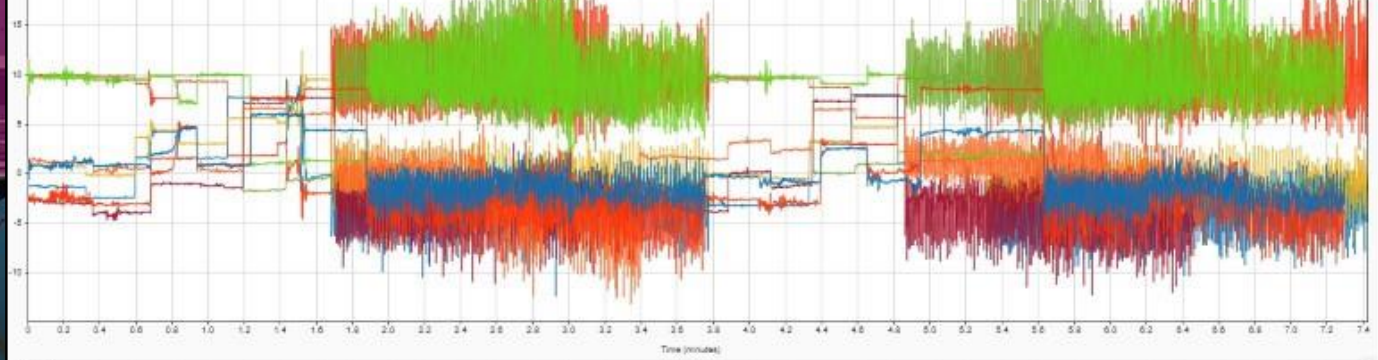
Our Expertise

From onboarding and implementation to solving advanced engineering challenges

7월 7일	7월 14일	7월 21일	7월 22일	7월 28일
영상 분석을 위한 딥러닝	신호처리를 위한 머신 러닝과 딥러닝	딥러닝 프로젝트를 위한 데이터 준비 기법	MATLAB Deep Learning Day 2021	MATLAB으로 시작하는 강화학습



<https://bit.ly/3hfSm24>
오늘 등록하세요!



감사합니다