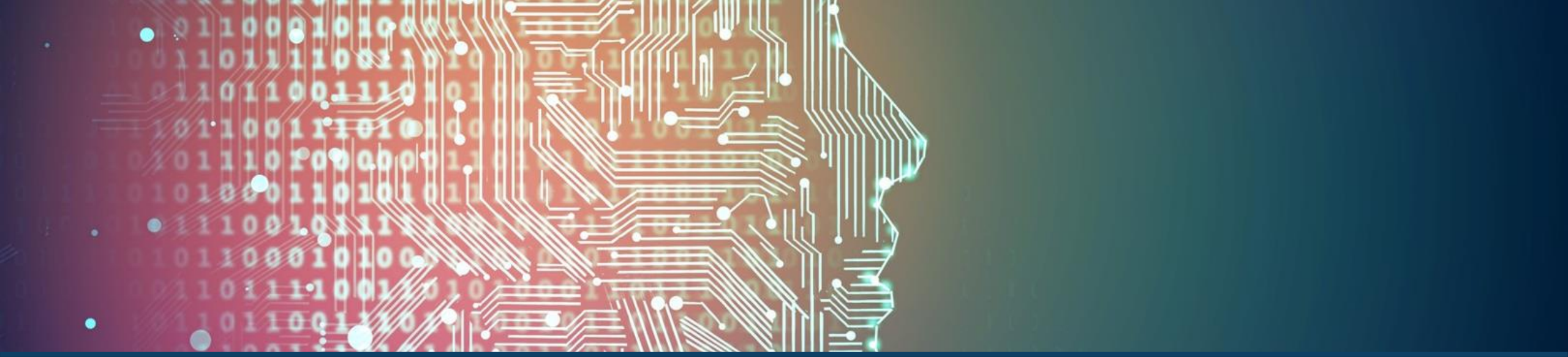


MATLAB과 함께하는 딥러닝 4주 완성 부트캠프

송완빈 과장
Application Engineer @ MathWorks
wsong@mathworks.com

MathWorks®

세션3. 딥러닝 프로젝트를 위한 데이터 준비 기법

MATLAB과 함께하는 딥러닝 4주 완성 부트캠프

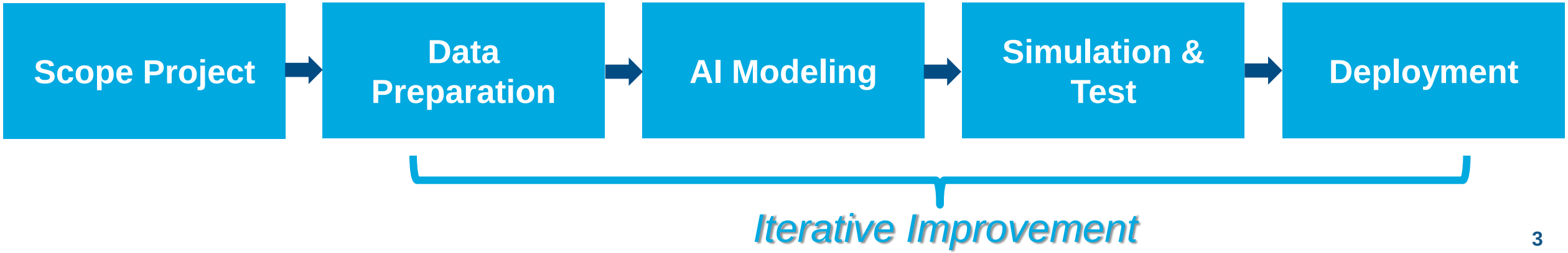
송완빈 과장
Application Engineer @ MathWorks
wsong@mathworks.com

MathWorks®
Accelerating the pace of engineering and science

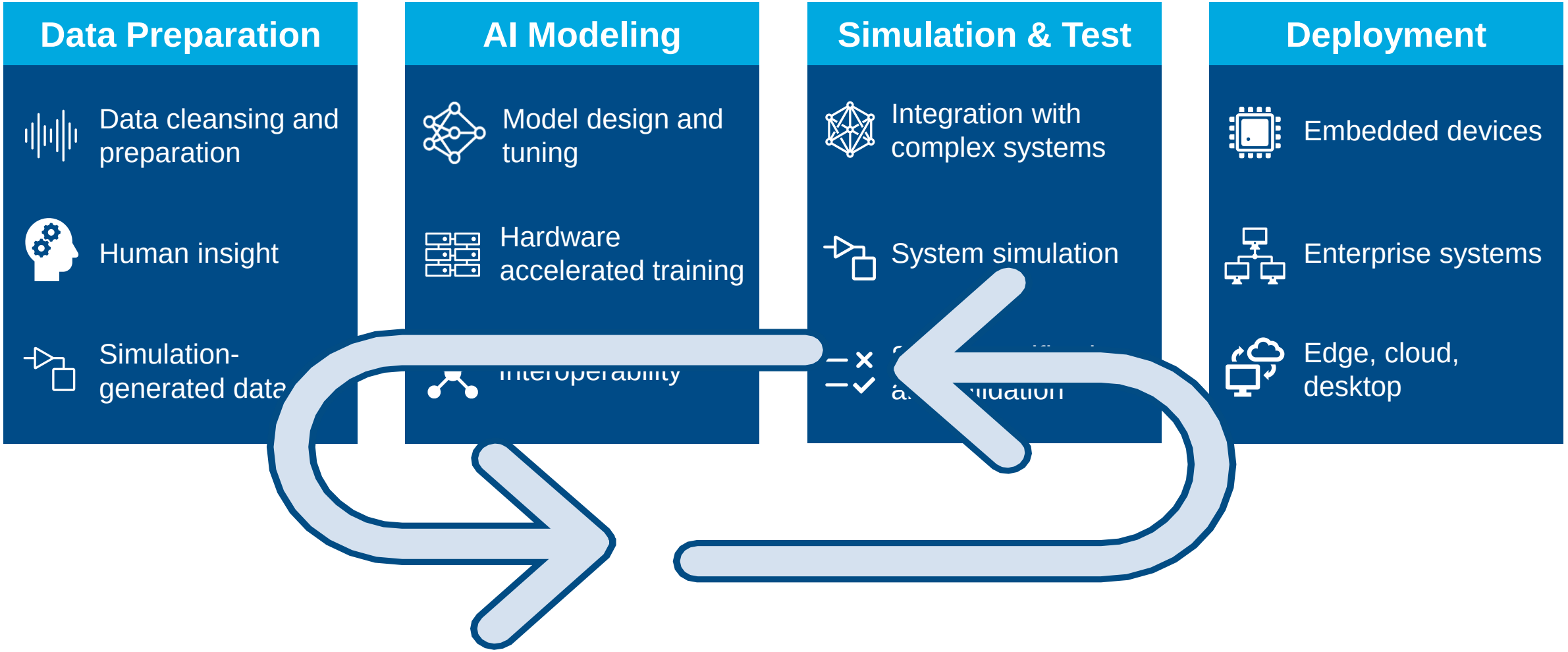
Software System Design

- Traditional Software

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



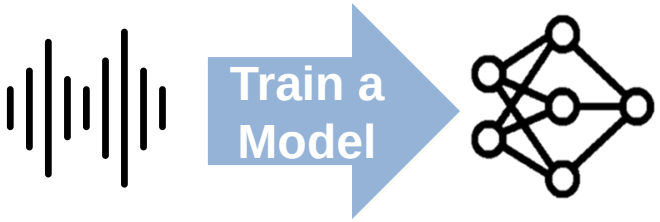
AI-driven system design

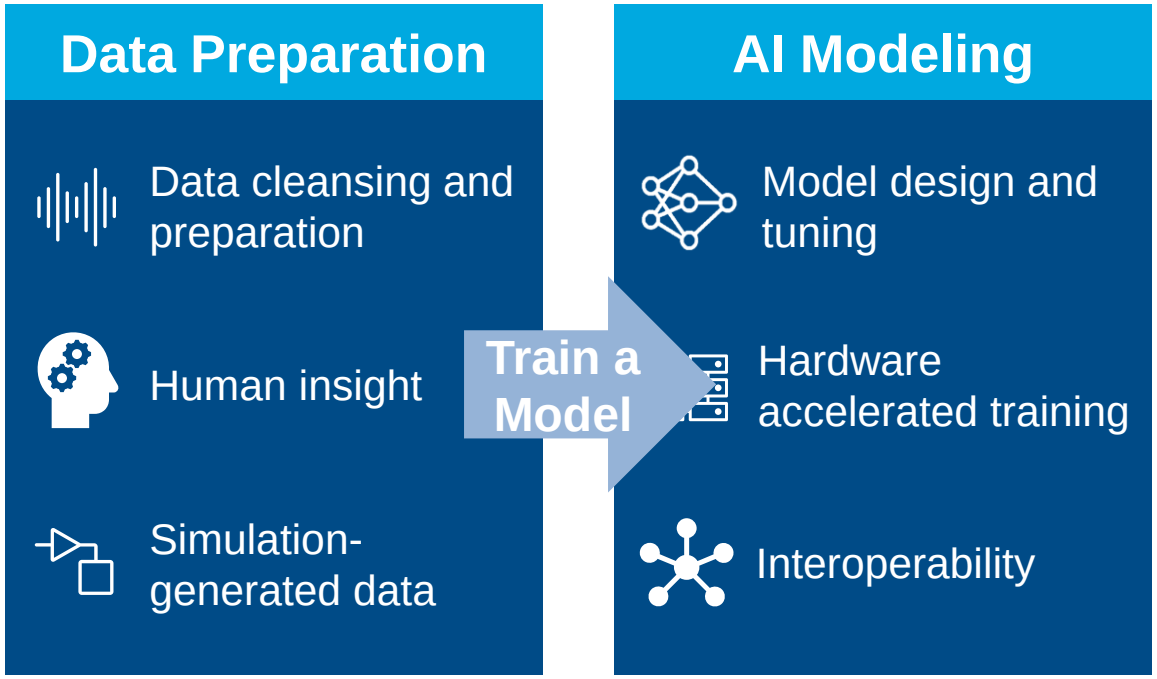


Iterative refinement for system improvement

4

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.

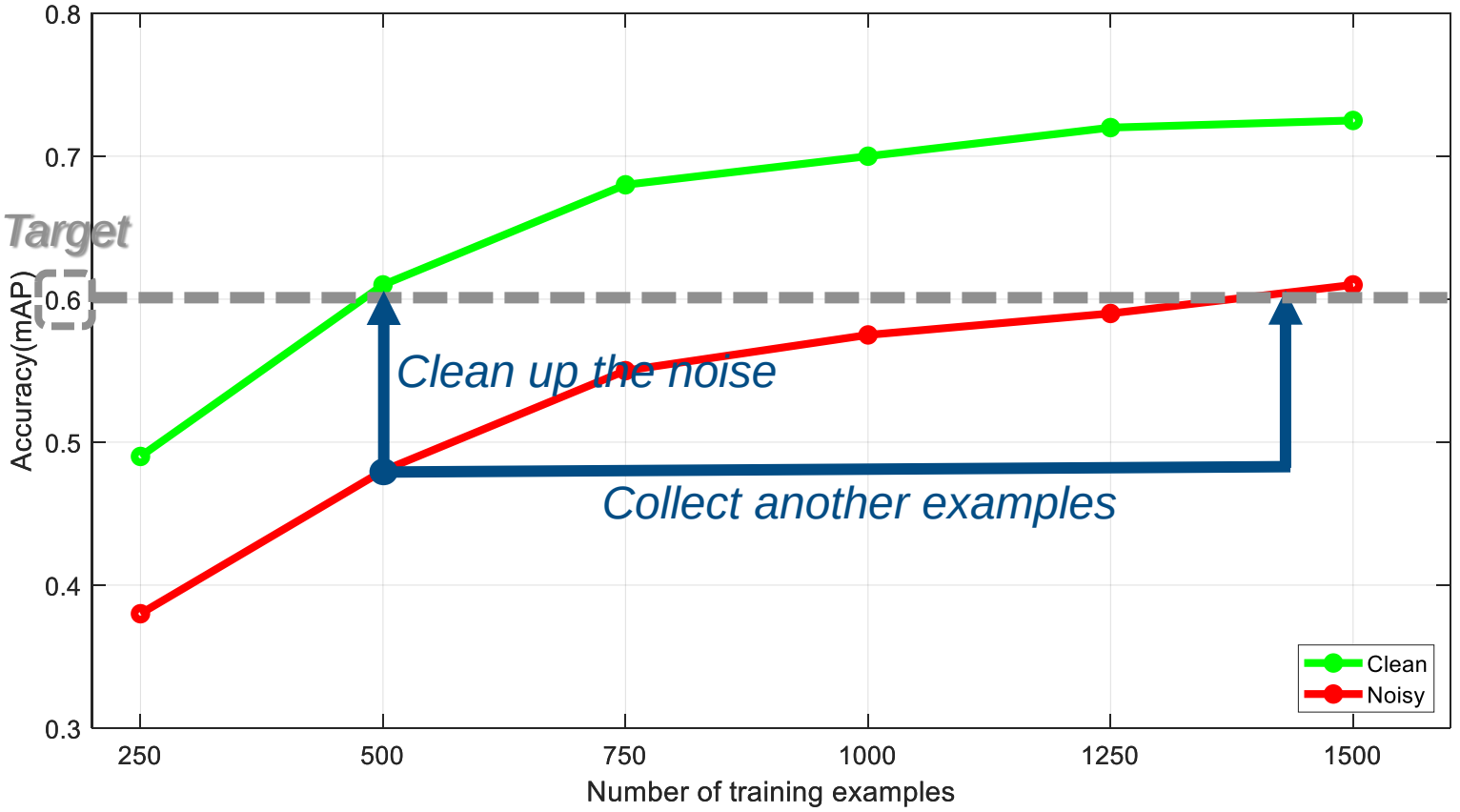
Data preparation represents most of your AI effort…

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

Data Preparation

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

Example : Clean vs. Noisy data



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

6

Data preparation represents most of your AI effort

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

Data Preparation

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

Example : Clean vs. Noisy data

Big Data to Good Data

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

7

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

8

Synthetic Data Generation and Augmentation to deal with less data

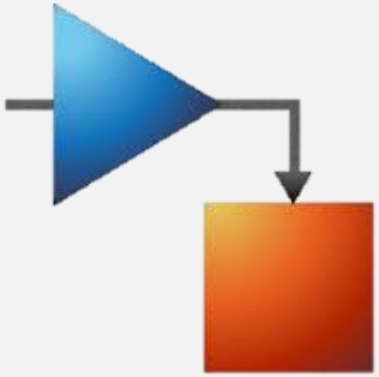
Simulate your system for synthetic data generation

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

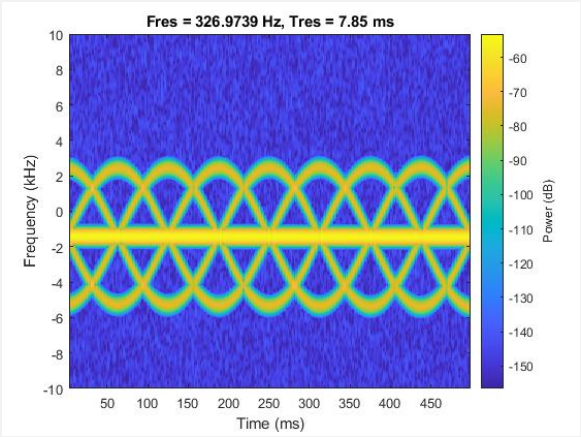
Data Preparation

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

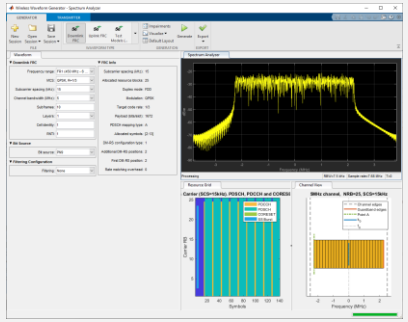
Simulate data using models and deep learning



Generate Radar Returns

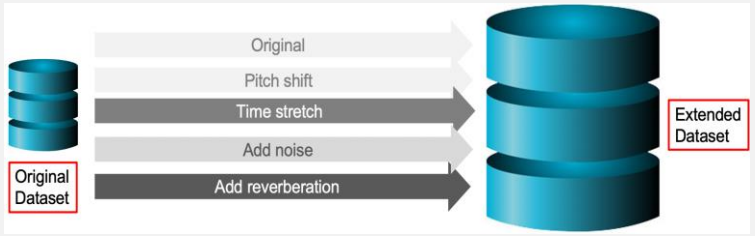


Generate wireless waveforms



Generate Audio Data

```
text2speech
```

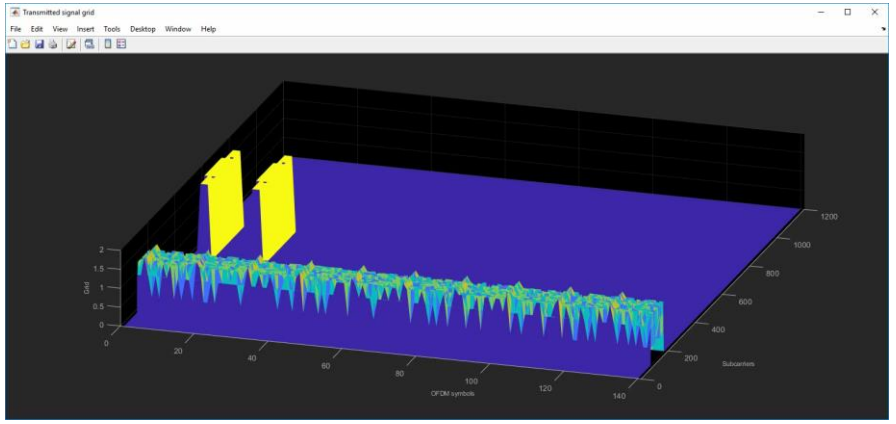


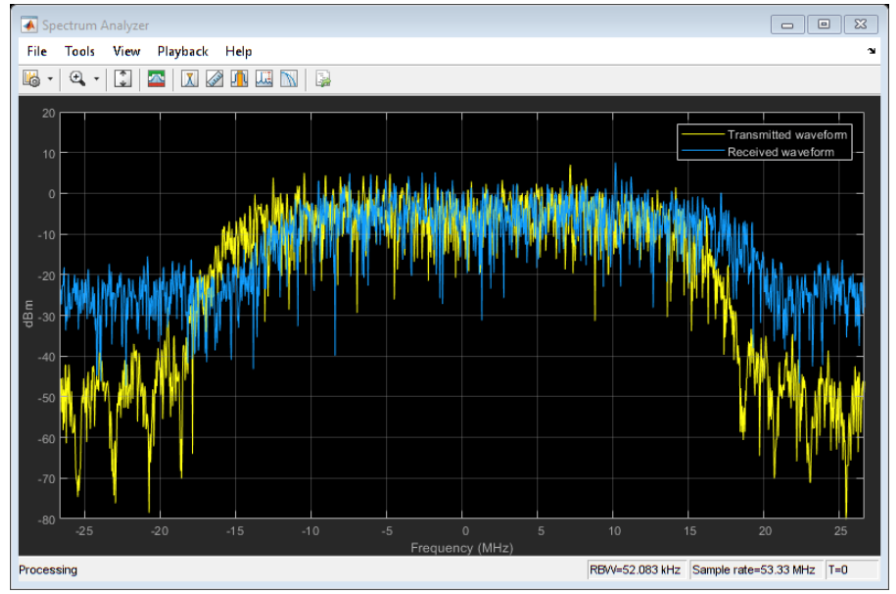
9

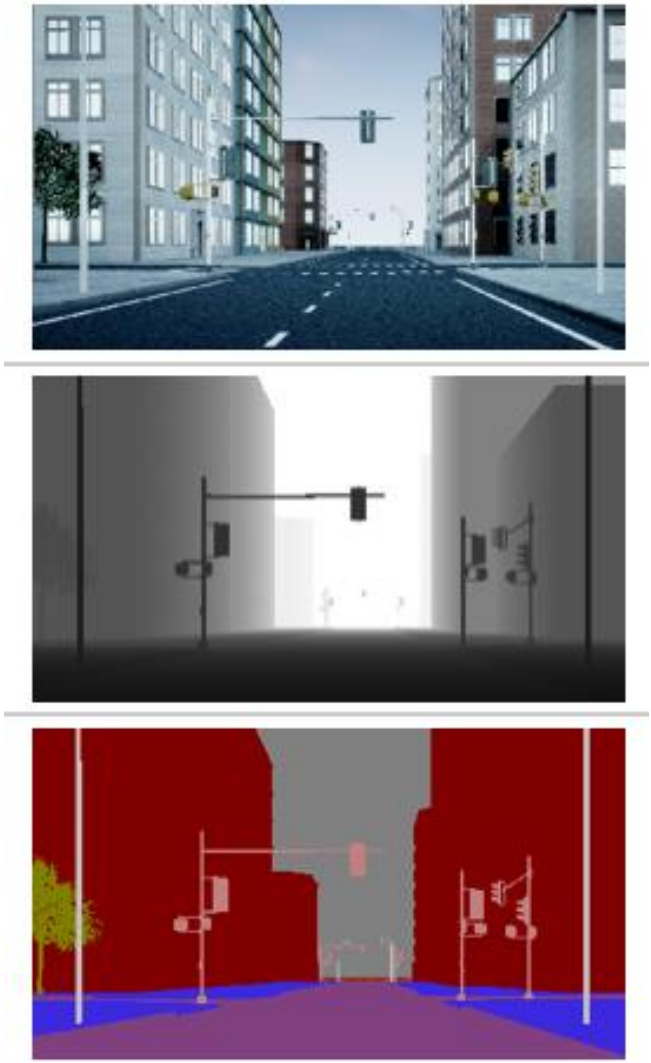
Synthetic Data Generation and Augmentation to deal with less data

Simulate your system for synthetic data generation

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





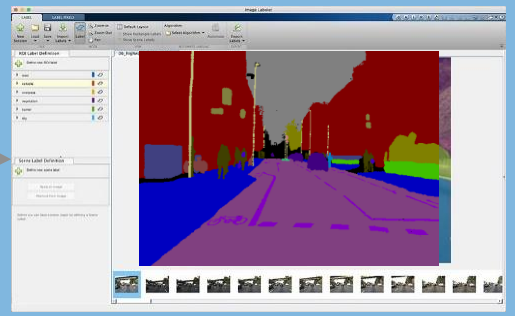


Data Preparation

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

One example of leveraging simulation for data synthesis

Traditional deep learning workflow



Record

Label

Simulation-based workflow

Simulate

Auto-label

AI model

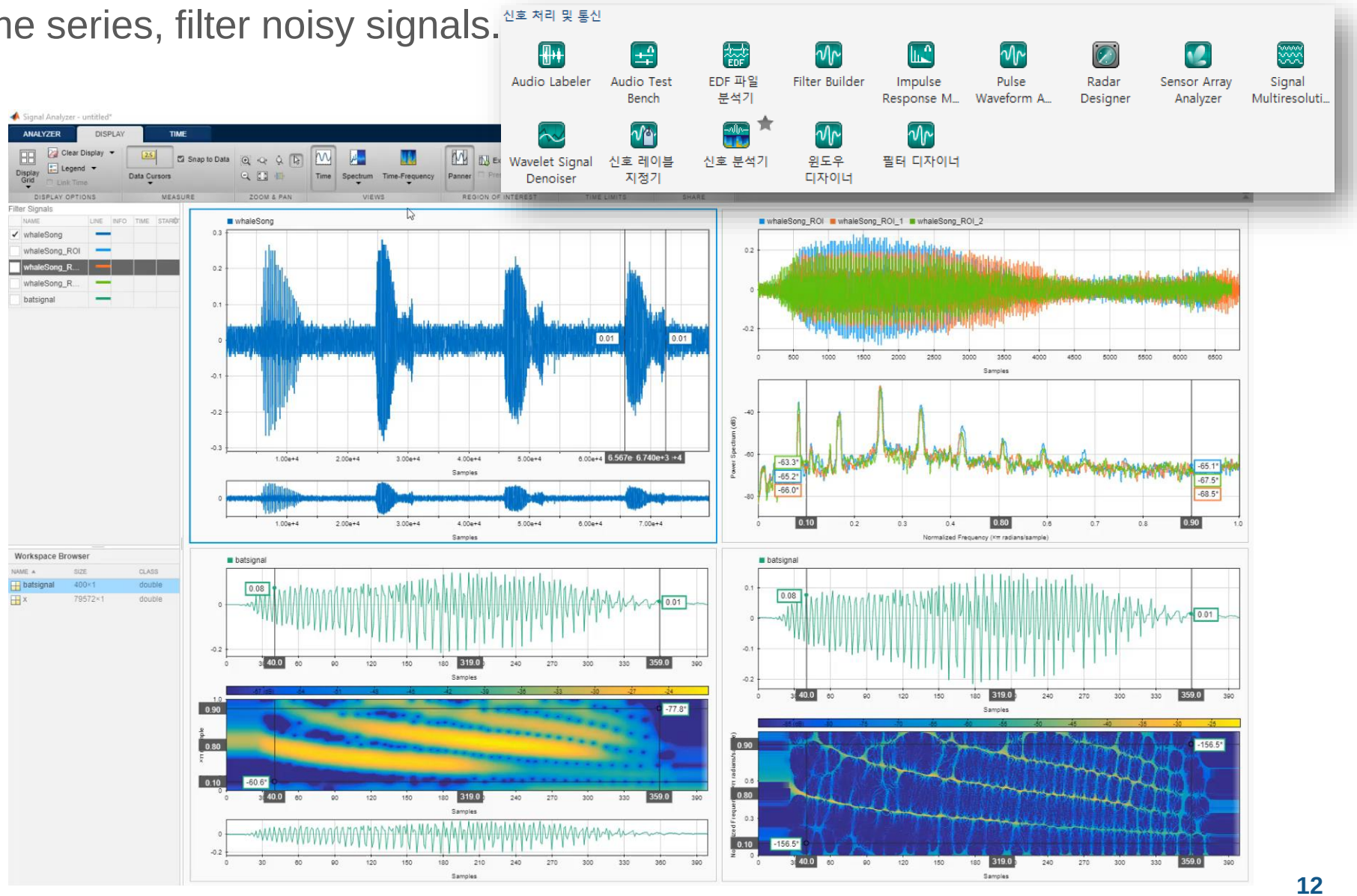
Spend less time preprocessing and labeling data

Synchronize disparate time series, filter noisy signals.

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

Data Preparation

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



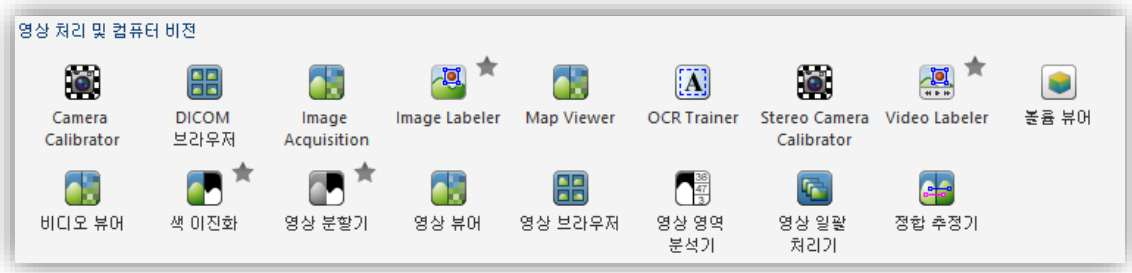
12

Spend less time preprocessing and labeling data

Preprocess image data using built-in Apps

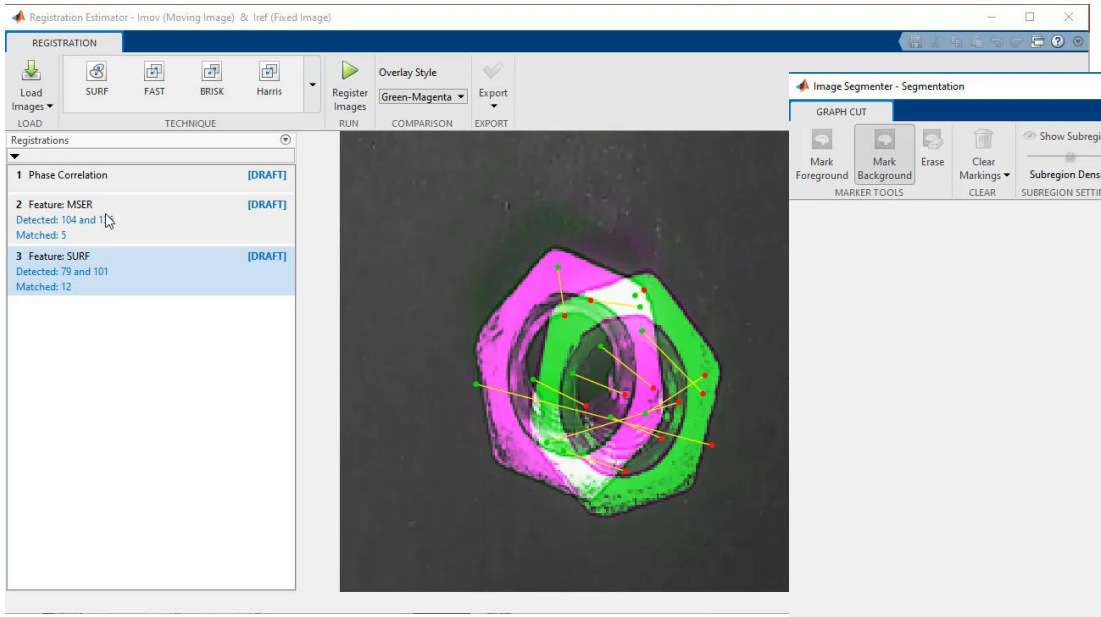
Data

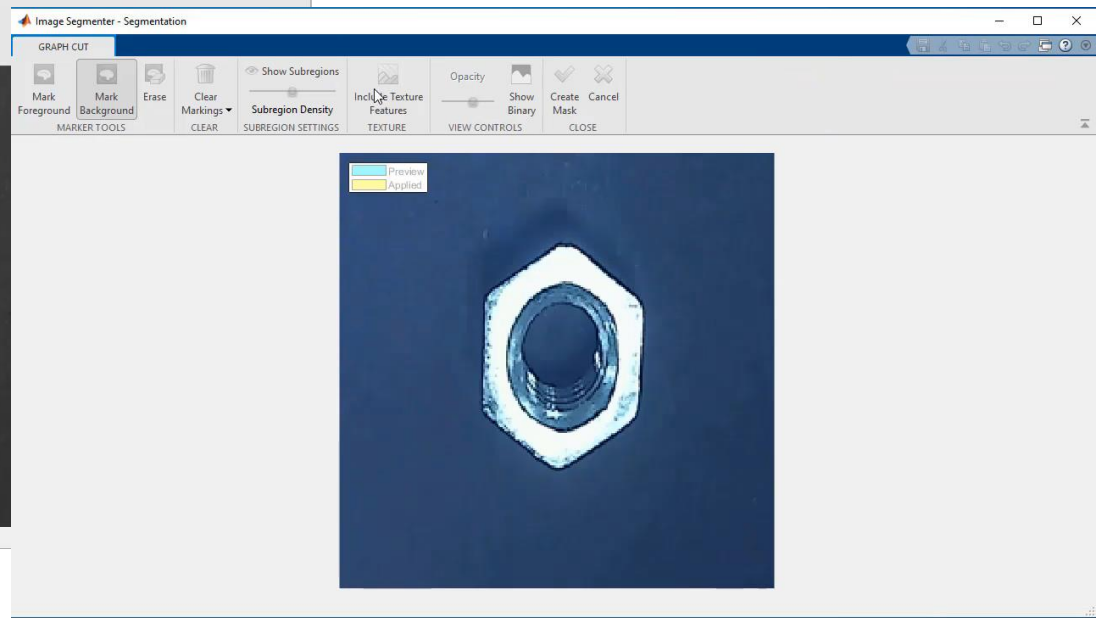
(**inputs x**, labels y)



Data Preparation

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






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

13

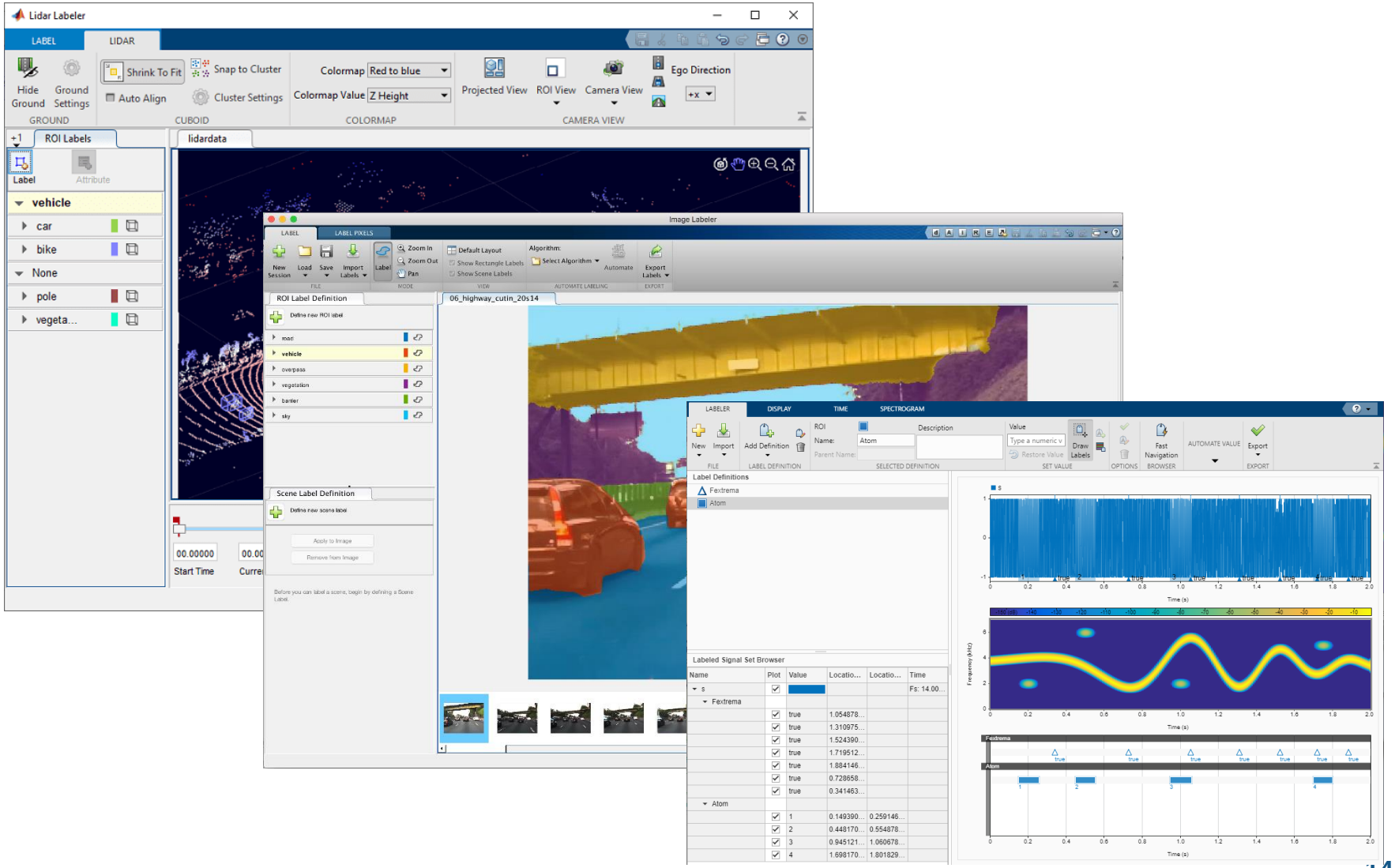
Spend less time preprocessing and labeling data

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

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

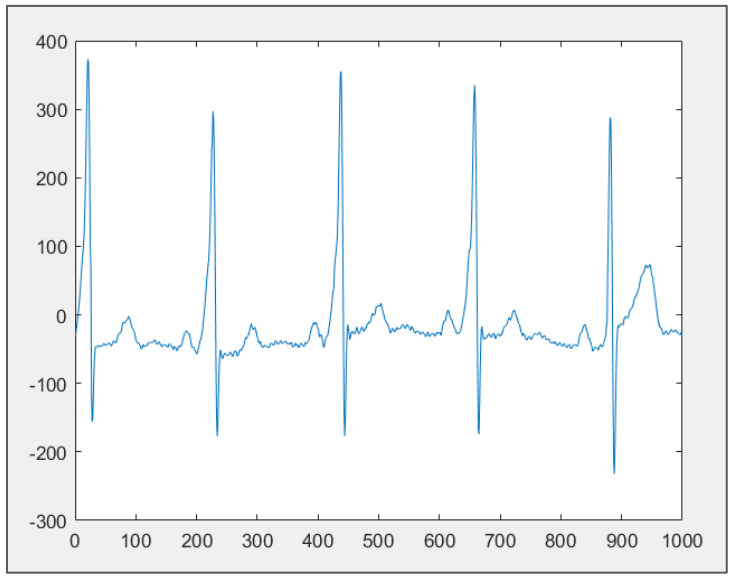
Data Preparation

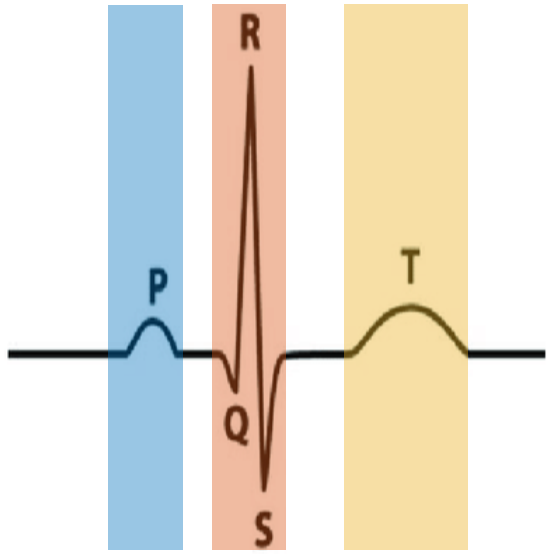
- Data cleansing and preparation
- Human insight
- Simulation-generated data

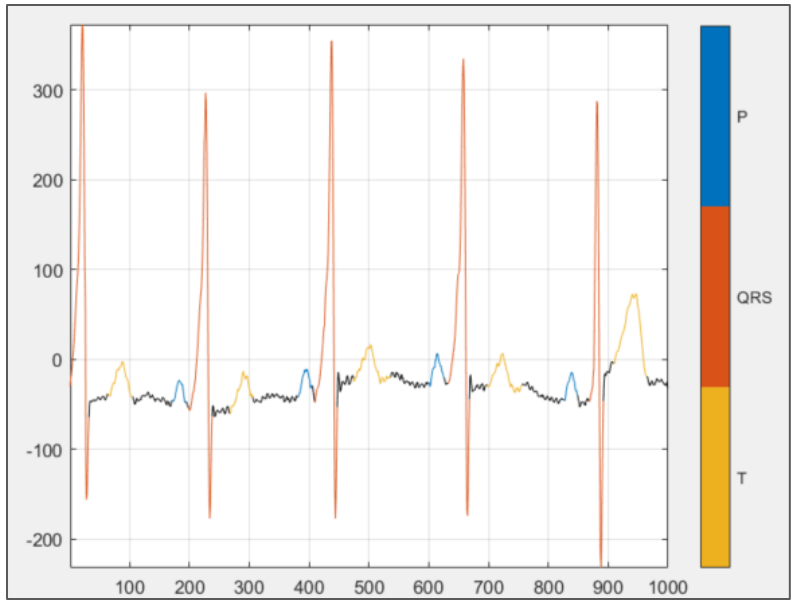


14

Example1 : Labeling key regions of ECG waveforms

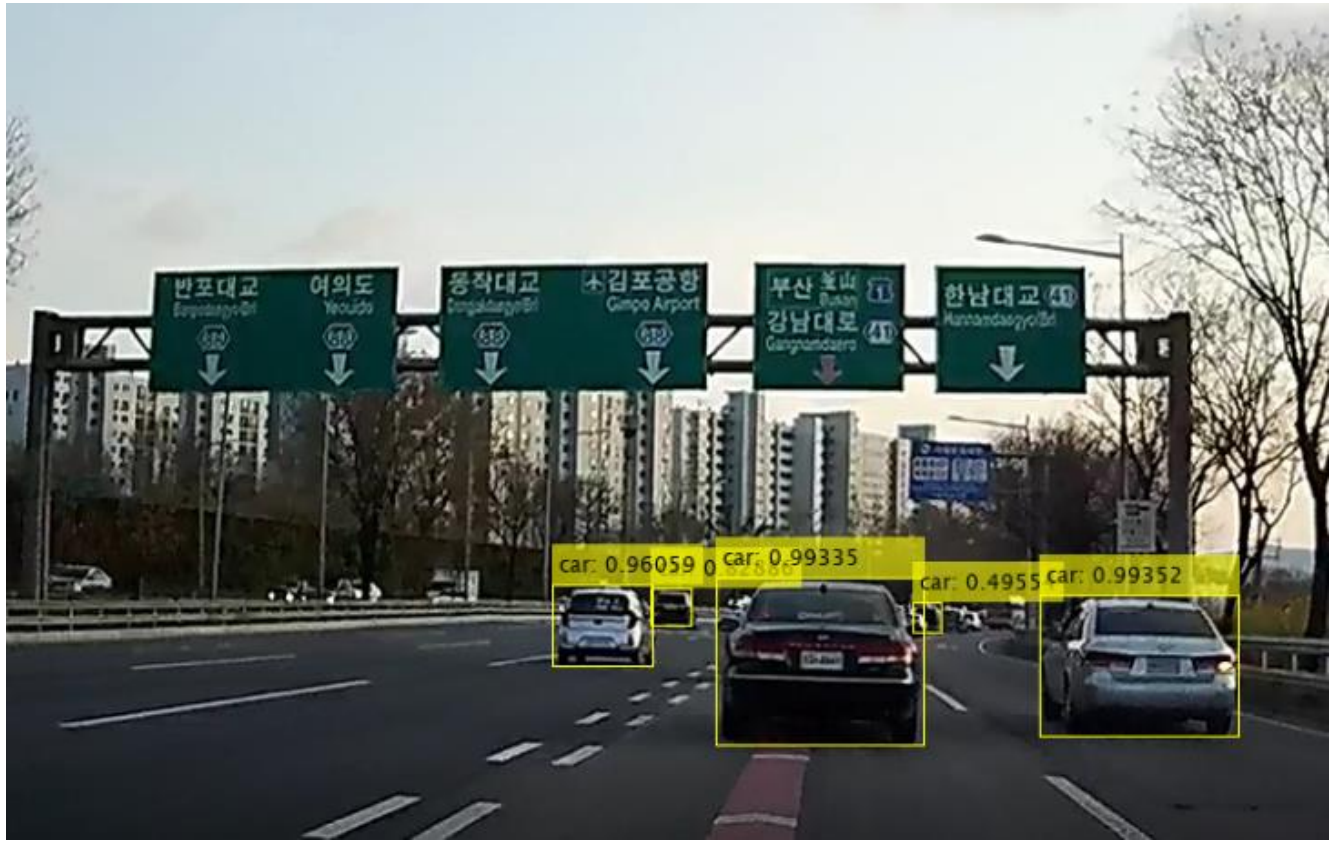






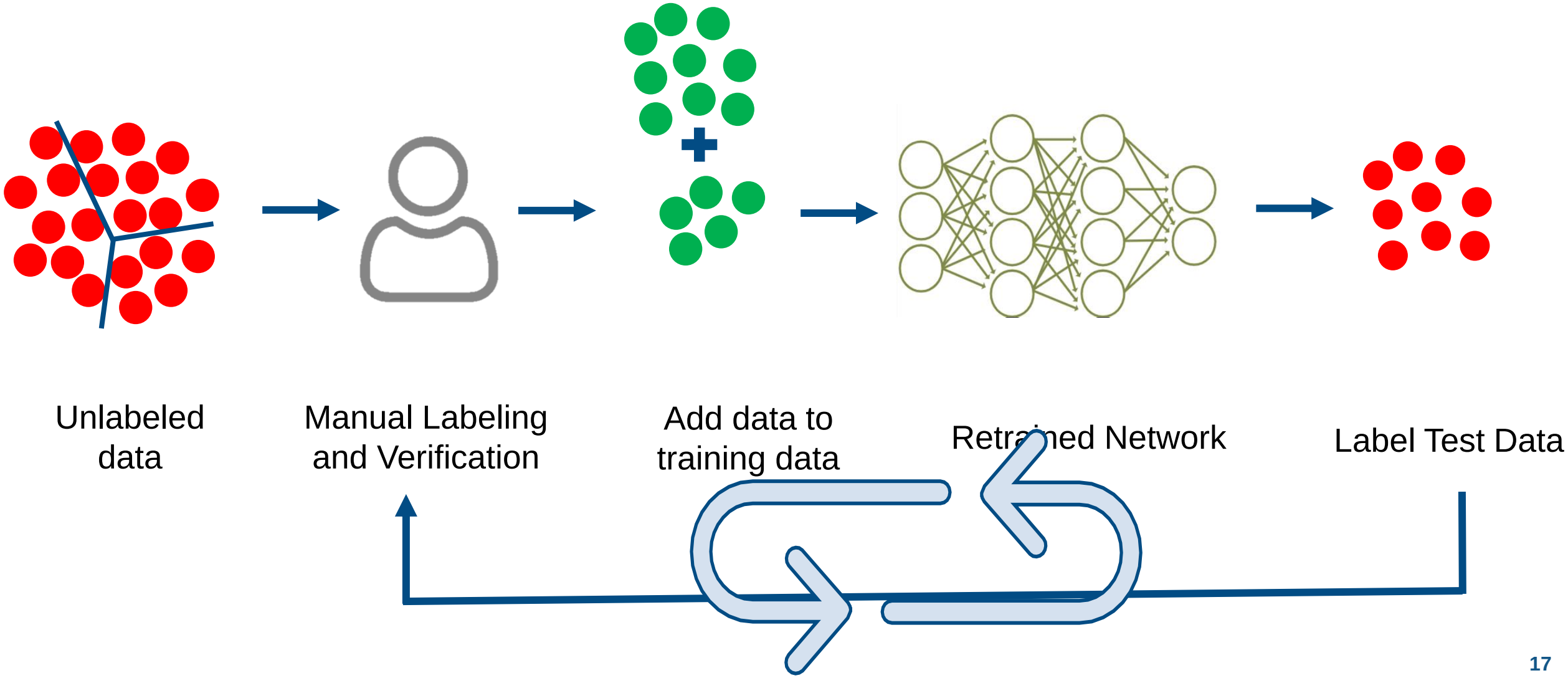
- ECG signals contain P,QRS and T waves.
- Identifying these can help with diagnosis and classification.
- Dataset contains 210 ECG signals ,~ 15 minutes long, labeled by cardiologist

Example2 : Labeling vehicles for object detection



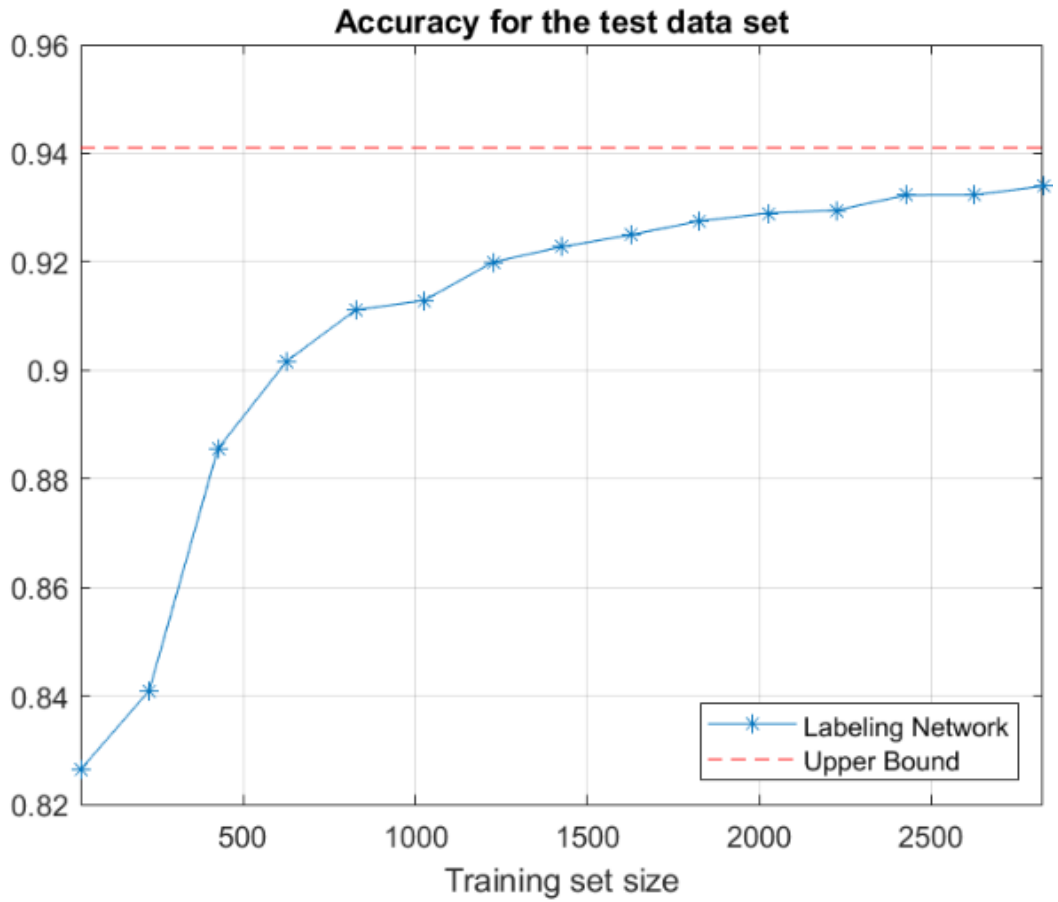
- Video data from my black box
- Wants to label cars in every frames of the video
- Do not have experiences on custom algorithm to detect vehicle
- Attended MATLAB Deep Learning Bootcamp #1

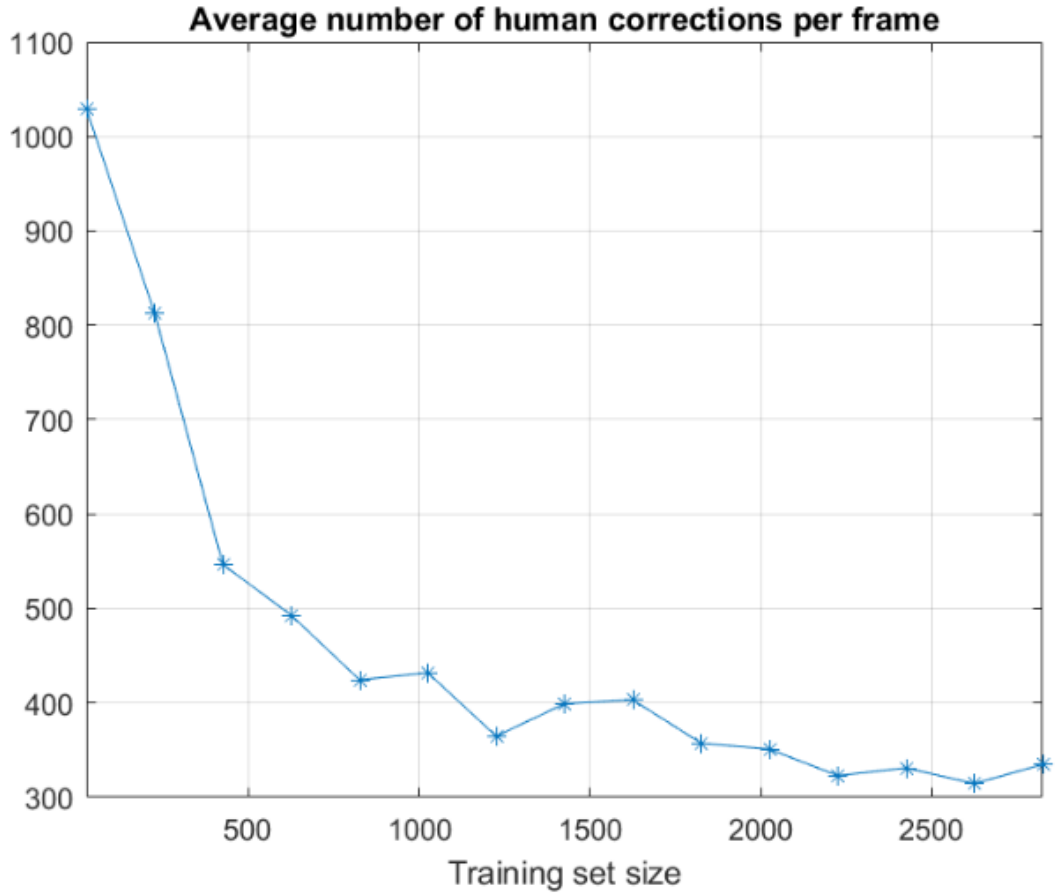
Establish iterative labeling process with deep learning



Result of iterative learning process

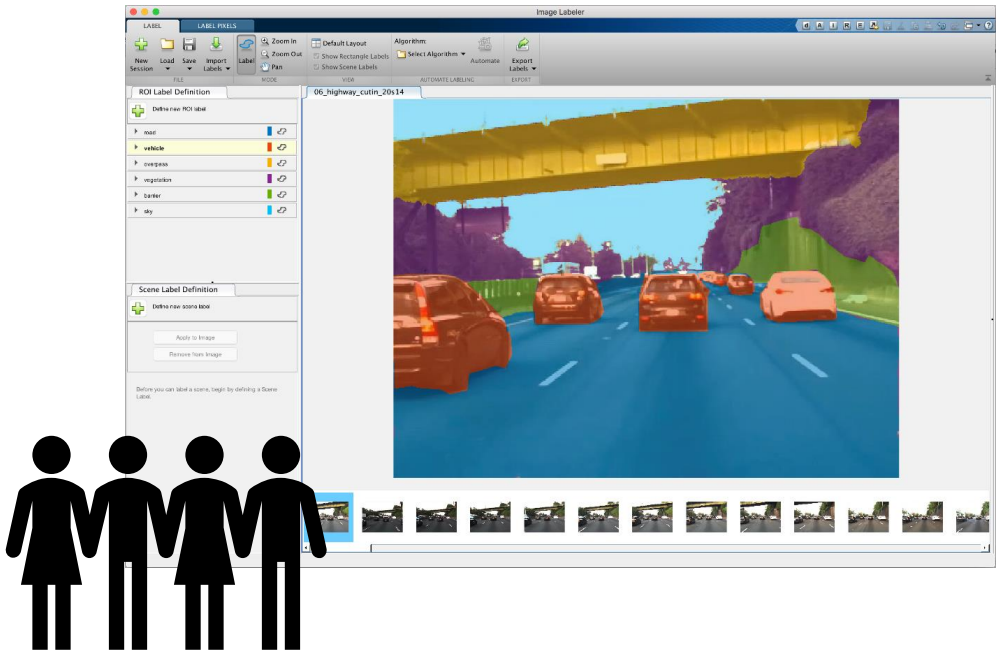
- Iteratively the accuracy increases, needs for human correction decreases.

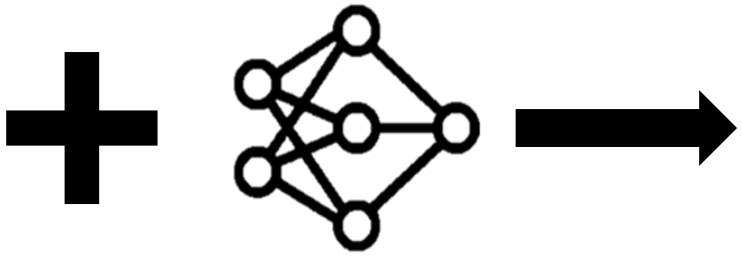


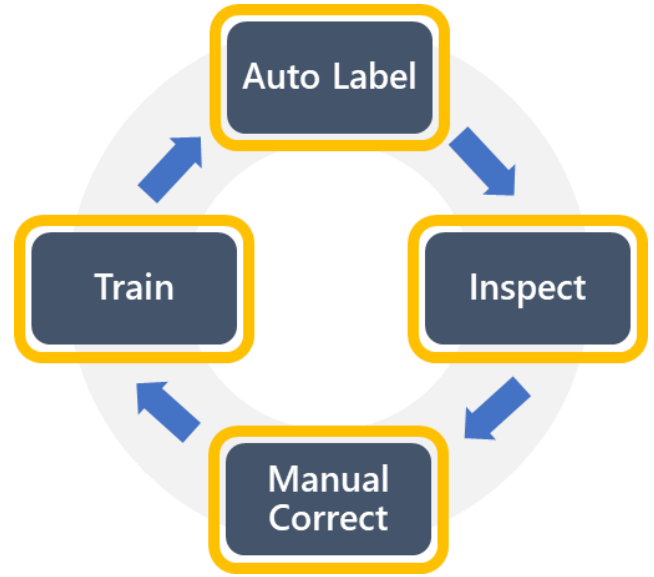


Strategy for automated labeling and iterative learning

- 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

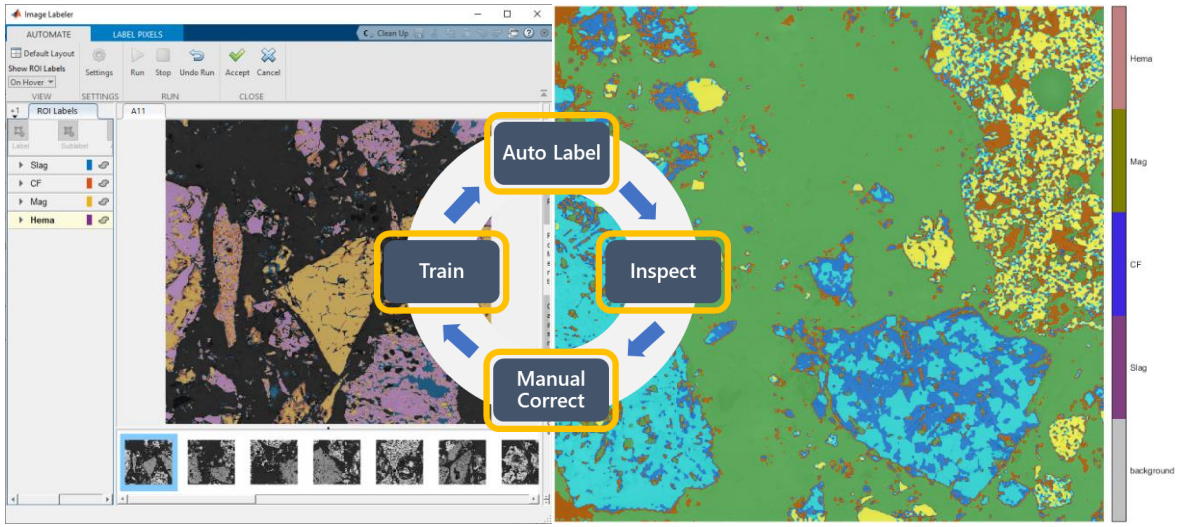
HYUNDAI STEEL – *HYUNDAI MOTOR GROUP*

> “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.”

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:

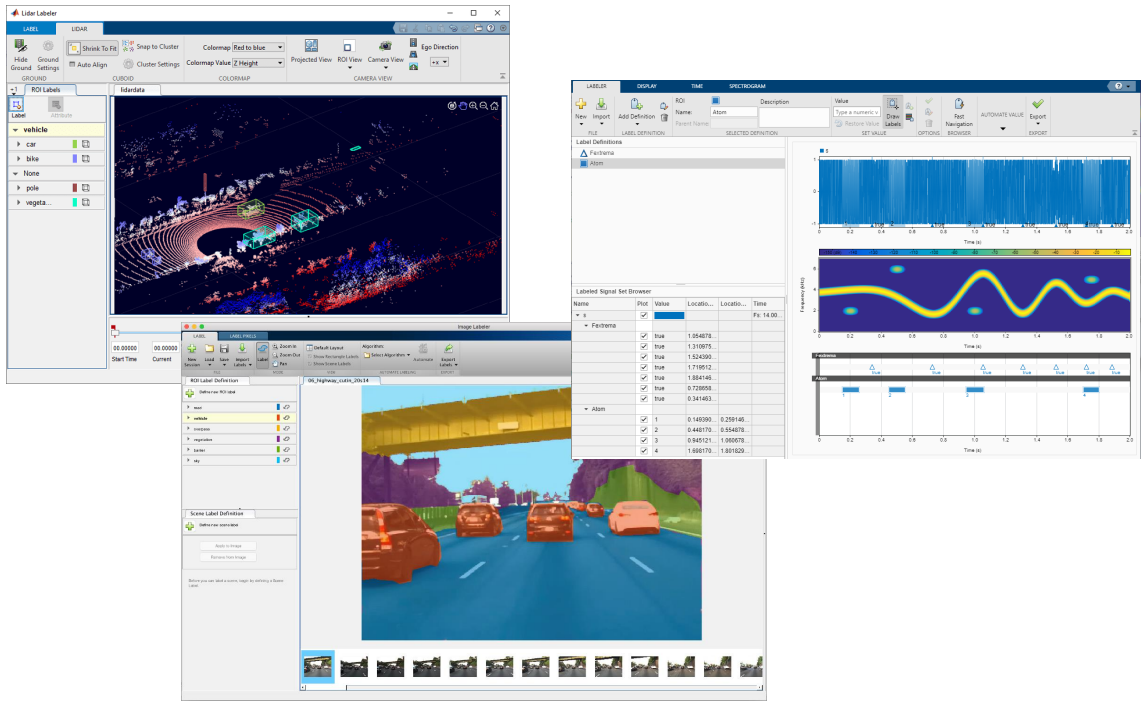
- Leveraged high-level APIs for applying image preprocessing techniques
- Interactive Apps enabled non-AI expert can do deep learning project
- Golden reference examples for quick prototyping
- Leverage MathWorks support to maximize the benefits of using MATLAB

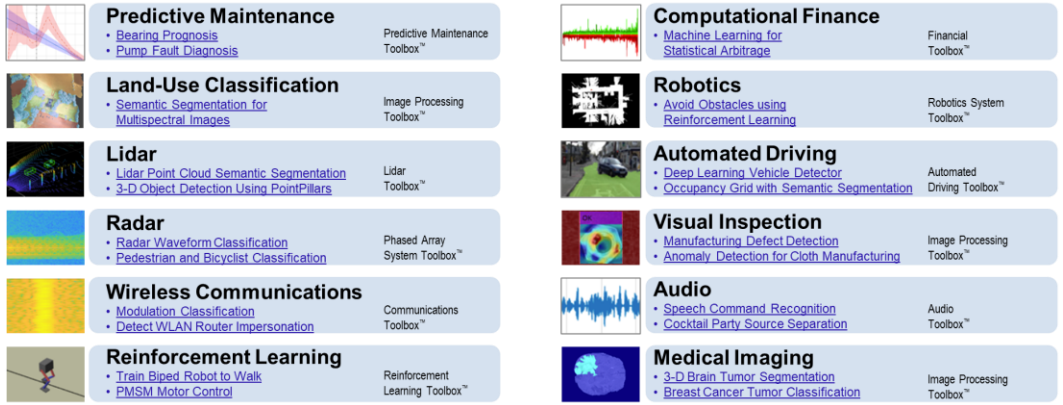
Explore MATLAB for deep learning

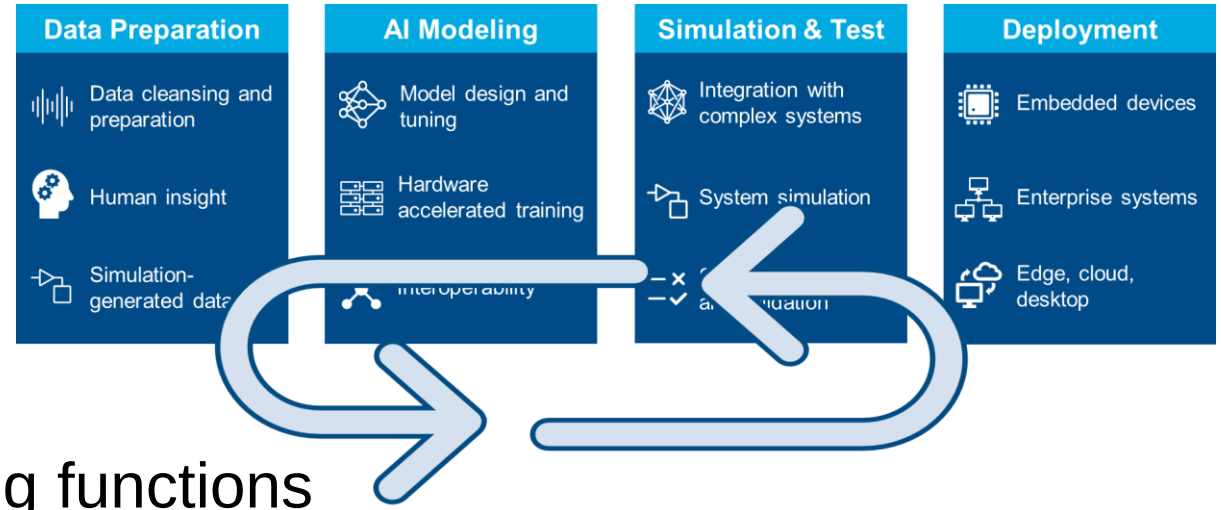
Semantic Segmentation Overview

Watch video

Build Systematic Data Management Process







- 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

딥러닝 부트캠프 응용편			
딥러닝 기반 이상탐지 기법	라이다 및 레이다를 위한 딥러닝	무선통신을 위한 딥러닝	바이오, 의료분야를 위한 딥러닝

https://bit.ly/3hfSm24
오늘 등록하세요!

MathWorks


70


Time (minutes)
Activity
Standing
Sitting
Walking
Lying Down
Climbing Stairs

감사합니다