Computer Vision Case Study

Suggesting: Pose detection and estimation

Abdul Rehman

Contact email: **abdulrehman786302@gmail.com**

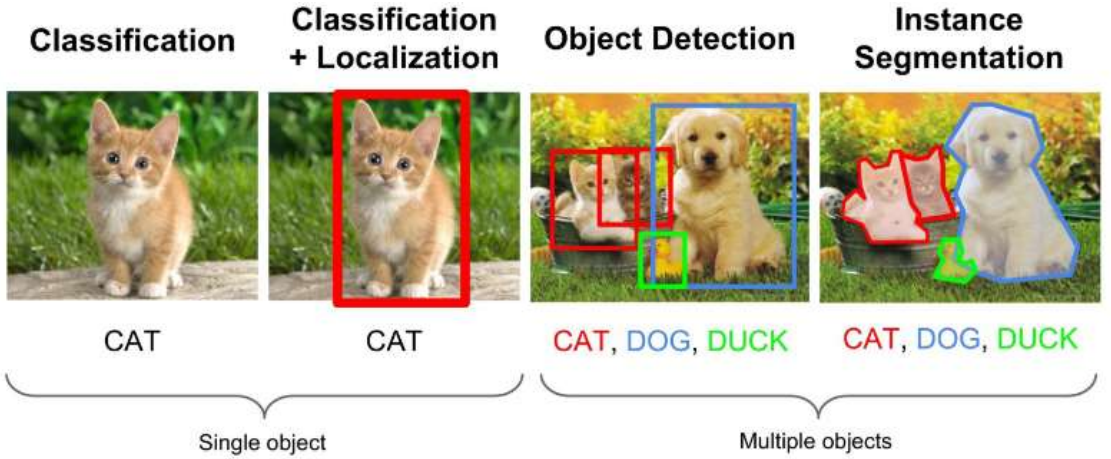
Abdul Rehman

Abdul Rehman earned a master's degree in multimedia networking from Telecom Paris Tech, France in 2020. He is now pursuing a doctorate in cyber security in multimedia at IMT Atlantique. The PhD is funded by IRT b<>com, a private research center in France.

His research focuses on the intersections of multimedia processing, AI, and deep learning.

1. Generic Setup Analysis

- Ideal Computer Vision Techniques:
  - Classification
  - Object detection
  - Segmentation
- Hardware Setup:
  - RBG 2D or 3D camera
  - ROS
  - Robot and tools bank
  - Operating system for vision



Camera → **Object Detection** → **Position or pose** → **Order & Number Operations** → **Remove Components**

1.1 Comparison Vision Techniques:

- **Industry Deep Learning Advantages:**
  - 360p cameras (UAVs, UGVs)
    - Spherical image complicates tradition CV methods.
  - 3D Vision (Classification, Geometry and Graphics)
    - Lots of development with 3D neural networks:
      - Computational complexity increases
    - 3D traditional CV even more difficult:
      - Extra dimensions increase uncertainty (e.g., occlusions, camera angles, etc.).

Traditional	Deep Learning
HOG(Histogram Oriented Gradient), SURF (Speeded Up Robust Features), Edge Detection and Motion Estimation	CNN Segmentation networks YOLO(You Only Look Once)

Differences	Deep Learning	Traditional
Manual feature extraction	No	Yes
Computationally-heavy	Yes	No
Huge label dataset	Yes	No
Easy to deploy	No	Yes
Accuracy	Yes	No

1.2 Detection with Deep learning

- CNN(Convolutional Neural Network):
  - CNN are made up from neurons.
  - Each neuron has a specific weight and bias.
  - CNN-architecture contains several layers.
  - Primarily an input layer, an ouput layer, and hidden layers.
  - A convolutional layer, a pooling layer, and other normalization layers make up the hidden layer.
  - A convolution method is performed by the convolutional layer.
  - Convolution in object detection is a method used to generate an output function from a filter.
  - The pooling layer reduce the dimension of the input.

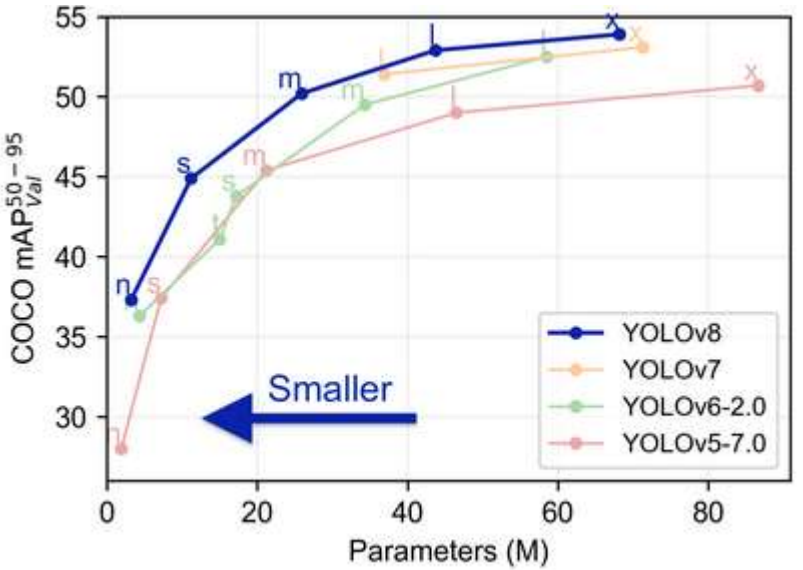
Region-Based Methods	Classification-Based Methods
R-CNN, Faster R-CNN	YOLO, SSD(Single Shot Detector)

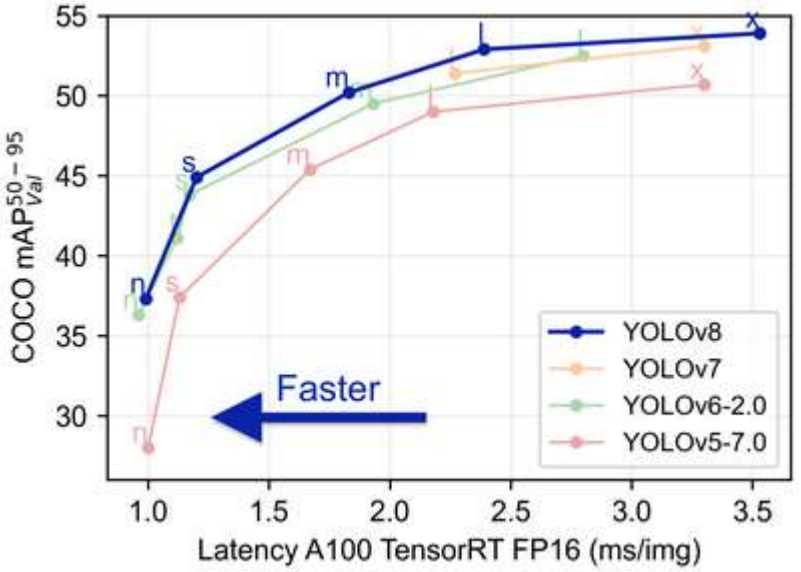
1.3 YOLO(You Only Look Once):

- YOLO(You Only Look Once):
  - One of the well known object detection algorithms
  - Speed and accurate
- Version of YOLO:

Version	Author	Release Date
V1	Joseph Redmon	Jun 2015
V2	Joseph Redmon and Ali Farhadi	Dec 2016
V3	Joseph Redmon and Ali Farhadi	Apr 2018
V4	Alexey Bochoknovskity	Apr 2020
V5	Glenn Jocher	May 2020
V6	Li et al.	Jun 2022
V7	WongKinYiu and (AlexeyAB)	Jul 2022
V8	Joseph et al.	Jan 2023

V8 is fastest.
V8 is light weight



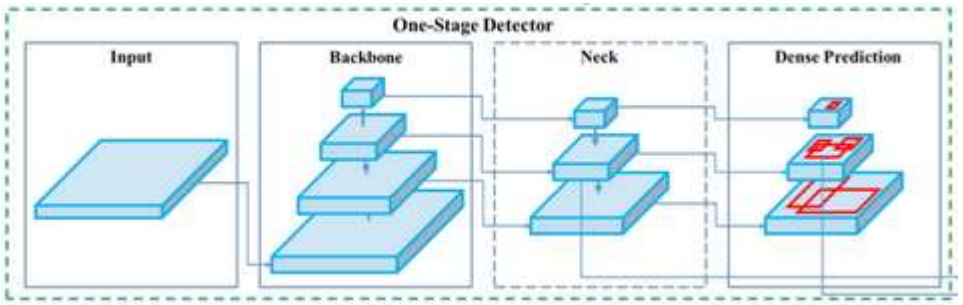


References

Lin, TY. *et al.* (2014). Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48.

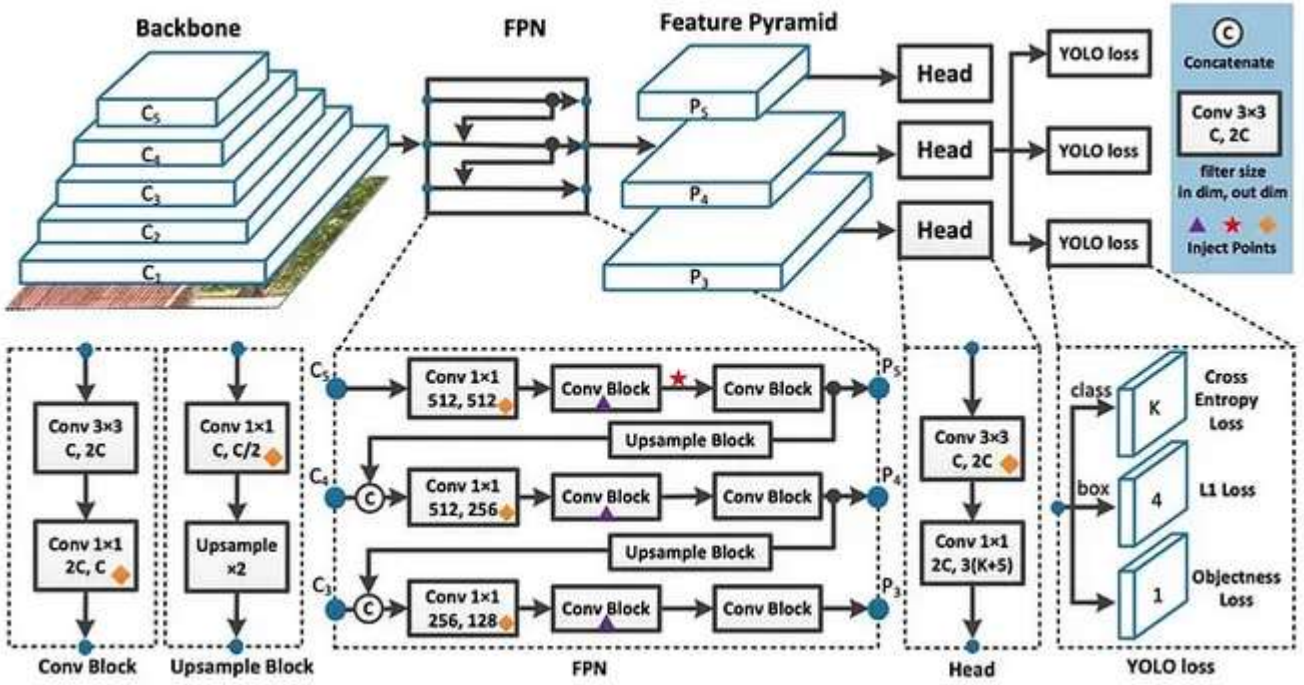
[74] YoloV5 Ultralytics. Tips for Best Training Reults. url: https://github.com/ultralytics/yolov5/wiki/Tips-for-Best-Training-Results. (accessed: 03.05.2024).

1.4 YOLO v8



- **The architecture** consists of a backbone, neck, and head.
- **The backbone,** is a pre-trained Convolutional Neural Network (CNN):
  - that extracts low, medium, and high-level feature maps from an input image.
- **The neck,** merges feature maps using path aggregation blocks like the Feature Pyramid Network (FPN).
- **The head**, classifying objects and predicting bounding boxes.

Anchor Free Detection:
YOLOv8 is an anchor-free model. This means it predicts directly the center of an object instead of the offset from a known anchor box.



2 Sensor Integration:

- **Zivid One+:**
  - A RGB-D camera which combines 2D images with the 3D technology, structured light.
  - 1920 x 1200 resolution, i.e. 2.3Mpixel
- **Induction sensor:**
  - Information about electrical conductivity
  - Electromagnetic field changes and signal fluctuations are measurable
- **Computer Hardware:**

Component	Type
Central Processing Unit (CPU)	AMD Ryzen 7 5800H (8 Cores)
GPU	NVIDIA GeForce RTX 3070 Mobile
Random Accessible Memory (RAM)	16GB (3200MHz)
Storage	M.2 SSD 1 TB

2.1 Deployment:

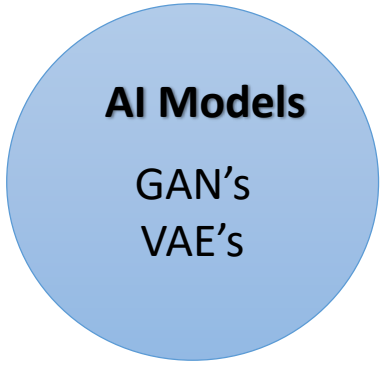
- **ROS:**
  - ROS is a meta-operating system for robots.
  - Number of independent nodes, each of which communicates with the other nodes using a publish/subscribe messaging model.
- **Realtime implementation:**
  - **Edge computing**
  - **Cloud computing**

	Edge Computing	Cloud Computing
Data Processing	Distributed among local devices and gateways	Centralized in remote data center
Costs	Less scalable expensive	Higher Scalability Lower operating costs
Latency	Lower latency	higher latency

2.2 Continuous Improvement

- **Current Challenges with Real Data**
- Companies often struggle to implement AI due to data-related challenges.
  - data regulations,
  - sensitivity, financial
  - financial Implications

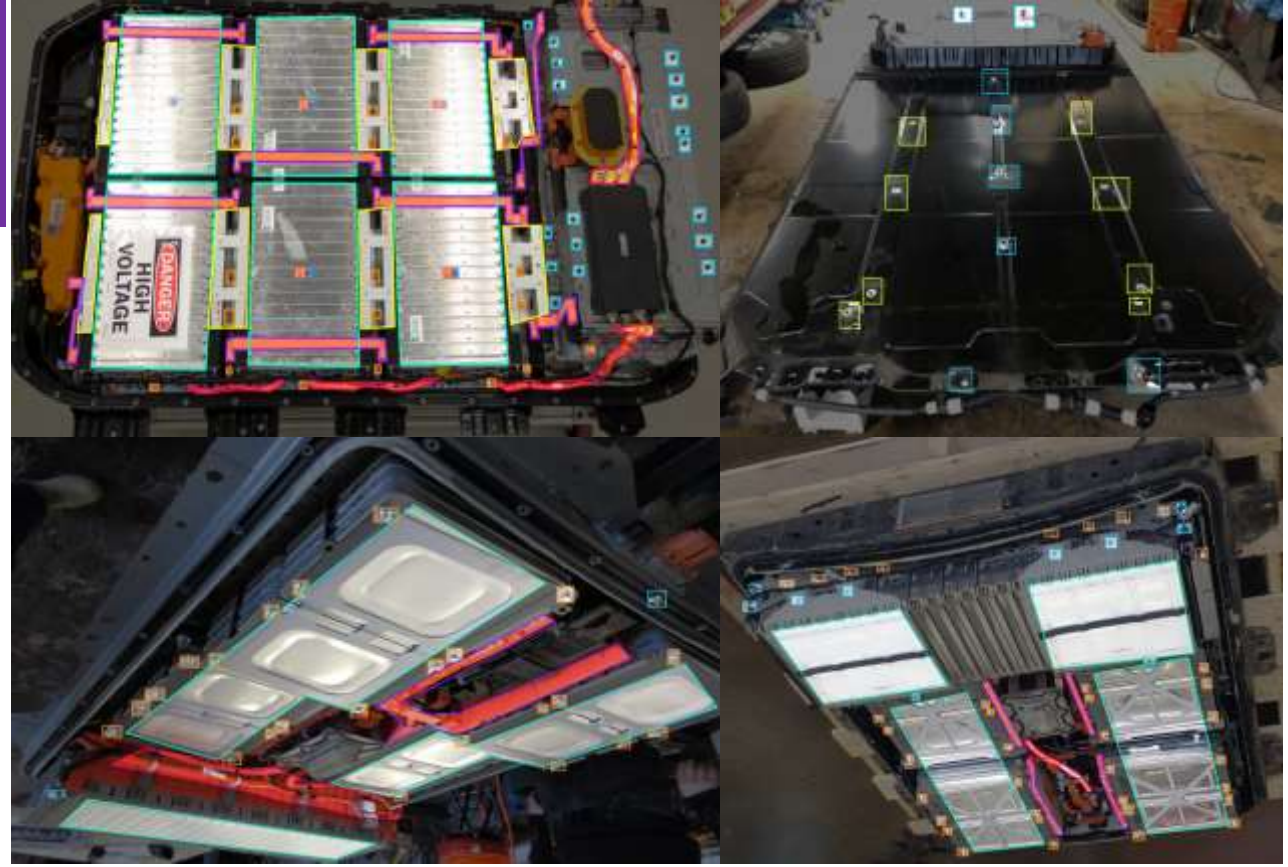
Area	Application of Generative AI	Benefits
Machine Learning	Enhancing datasets in limited or non-diverse scenarios.	Improves training of machine learning models by introducing diverse data points.
	Augmenting training datasets for model robustness.	Enhances model performance, especially in data-restricted situations.
Testing	Testing data-centric applications (data pipelines, algorithms, software).	Validates the performance and resilience of various system components.
	Creating specific test cases and scenarios.	Assists in assessing system performance and resilience in edge cases and outliers
Data Augmentation	Expanding training datasets in fields like computer vision.	Increases dataset size and diversity, aiding in model training.
	Generating additional data points.	Reduces overfitting, and enhances performance on new data



- **GANs**, a neural network that includes both a "generator" and a "discriminator."
  - Generator produces synthetic data that closely resembles real data,
  - Discriminator effectively differentiates between genuine and synthetic data.
- **VAEs** use an "encoder" and a "decoder" to generate synthetic data.
  - Encoder summarizes the characteristics and patterns of real-world data.
  - Decoder tries to convert that summary into a synthetic dataset that is very similar to real data.

3.1 Coding Challenge: EV Battery pack

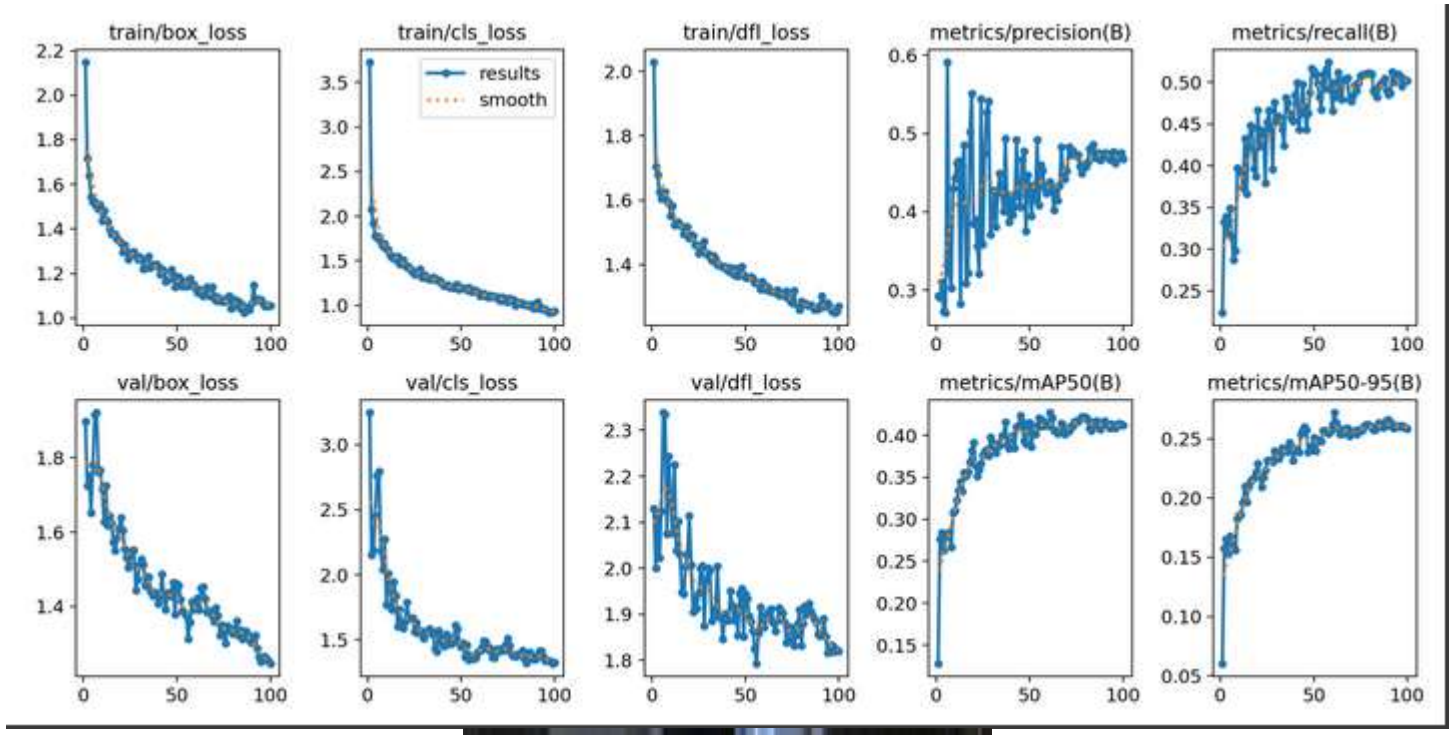
- Classes
  - Frame
  - Battery Module
  - Screw
  - Cable
  - Bars
- Augmentation:
  - Flip: Horizontal, Vertical
  - Shear: ±10° Horizontal, ±10° Vertical
  - Saturation: Between -25% and +25%
  - Brightness: Between -15% and +15%
  - Exposure: Between -10% and +10%
  - Blur: Up to 2.5px
- Split: 345 Images
  - Training: 87%, Validation: 10% and Testing: 3%

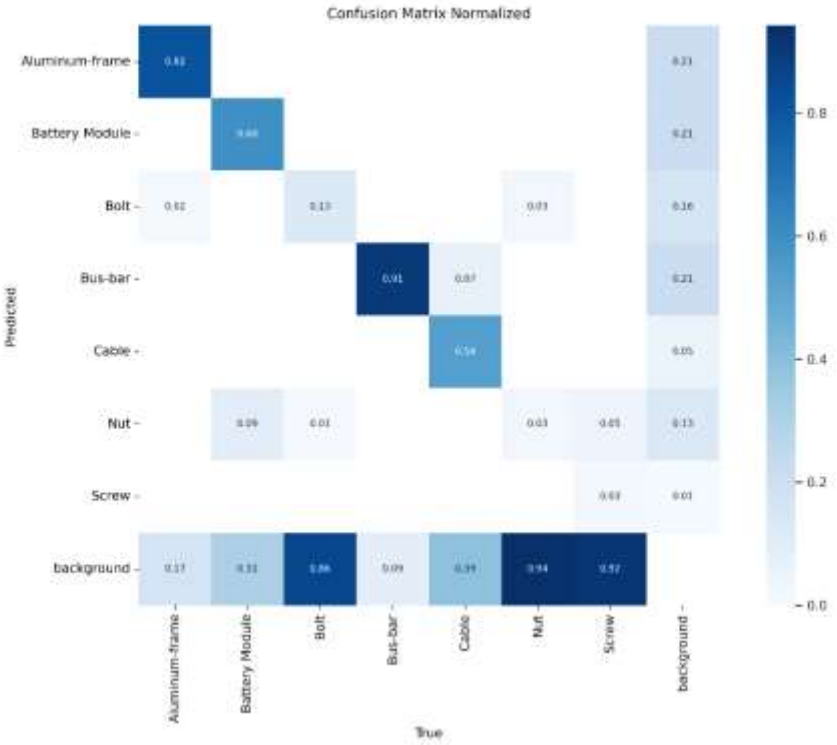


Reference:

url ={ https://universe.roboflow.com/mtech-project-ohj8a/ev-battery-pack }, journal = { Roboflow Universe }, publisher = { Roboflow }, year = { 2024 }, month = { mar }, note = { visited on 2024-03-13 },

3.2 Yolov8: Trianing & Testing





4. Conclusion

- **DL based object detection:** YOLO v8
  - Faster
  - Lighter
- **Generative AI and Synthetic Data:**
  - Explore GANs and VAEs
- **EV Battery pack dataset:**
  - Annotate
  - Augmentation
- **Train & test yolov8:**

Thank you !!
Please don’t hesitate to send an email for questions.

Contact email: **abdulrehman786302@gmail.com**