Loopr AI

Data Science Assignment

Problem Statement:

Develop a defect detection system for carbon fiber products, using supervised or unsupervised learning techniques. The goal of this assignment is to build a model that can accurately detect various classes of defects in carbon fiber such as holes, tears, stains, misalignment and other types.

Dataset is available here: https://www.kaggle.com/datasets/nexuswho/aitex-fabric-image-database

Your main tasks for this assignment include:

1. Data Exploration: Begin by exploring above dataset and select which class(es) of defects your model would identify (more is better, but we expect the model to identify at least 2 classes of defects)

2. Feature Engineering: Based on your analysis, identify relevant features that can help distinguish between defective and non-defective carbon fiber samples. Consider both image-based features (e.g., pixel intensity, texture) and additional features provided in the dataset.

3. Build a supervised or unsupervised learning model depending on what you think is the best approach to solve this problem with high accuracy

Supervised Learning: Model such as object detection, segmentation to classify the carbon fiber samples into different defect categories. Train the model using the labeled dataset and evaluate its performance using appropriate metrics (e.g., accuracy, precision, recall).

Unsupervised or semisupervised Learning: Utilize unsupervised or semi supervised learning techniques (such as transformers or encoders) to identify any hidden patterns or anomalies in the dataset. Explore whether these techniques can help identify new types of defects or detect outliers in the data.

4. Model Deployment/Evaluation: Finally, deploy or host the model on any public endpoint where we can run inference to evaluate the model. If it's not possible to host the model, then provide us instructions on how to build and run the model locally.

5. Communication: Present results of the model as though you're trying to convince a stakeholder that the model you've built is ready to be integrated into a production line. Build a brief presentation in your preferred form or software and include it in your Github repository.