Defect Spectrum: A Granular Look of Large-Scale Defect Datasets with Rich Semantics

Shuai Yang[1,2]*, **Zhifei Chen**[1]*, **Pengguang Chen**[3], **Xi Fang**[3], **Shu Liu**[3], **Yingcong Chen**[1,2,4]

[1]HKUST(GZ), [2]HKUST(GZ) - SmartMore Joint Lab, [3]SmartMore, [4]HKUST

Abstract

Defect inspection is paramount within the closed-loop manufacturing system. However, existing datasets for defect inspection often lack precision and semantic granularity required for practical applications. In this paper, we introduce the Defect Spectrum, a comprehensive benchmark that offers precise, semantic-abundant, and large-scale annotations for a wide range of industrial defects. Building on four key industrial benchmarks, our dataset refines existing annotations and introduces rich semantic details, distinguishing multiple defect types within a single image. Furthermore, we introduce Defect-Gen, a two-stage diffusion-based generator designed to create high-quality and diverse defective images, even when working with limited datasets. The synthetic images generated by Defect-Gen significantly enhance the efficacy of defect inspection models. Overall, The Defect Spectrum dataset demonstrates its potential in defect inspection research, offering a solid platform for testing and refining advanced models.
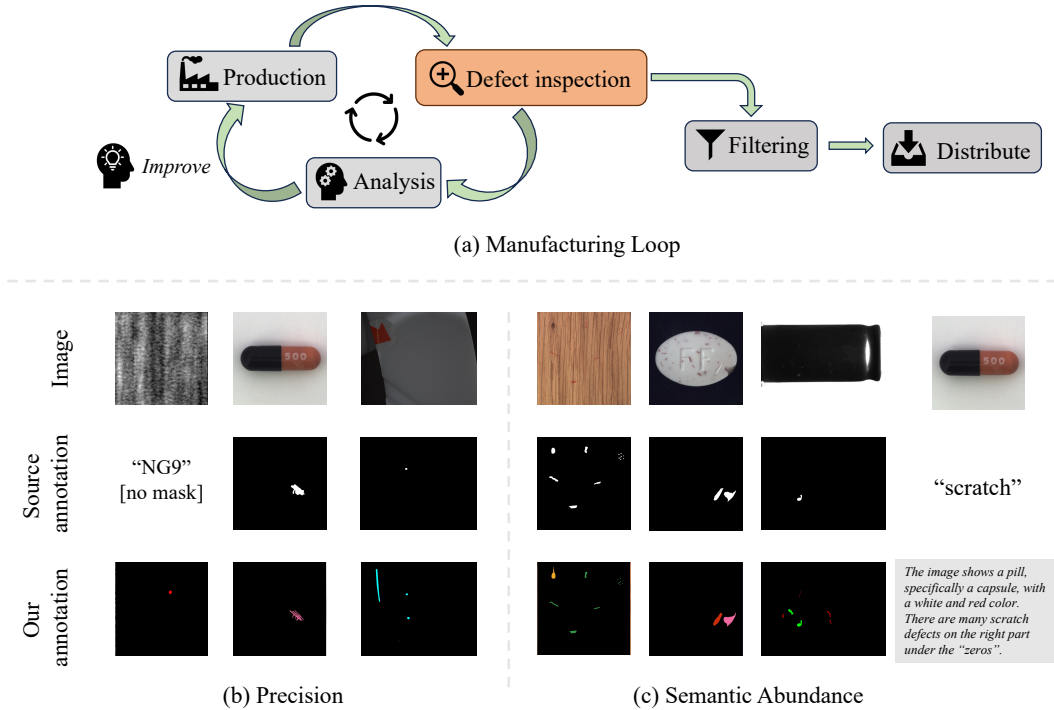
Figure 1: (a) The closed-loop system in industrial manufacturing. Defect inspection plays a pivotal role. (b, c) An overview of our improvements in annotations, in the aspect of precision and semantics abundance. **Best viewed in color.**

---

*These authors contributed equally to this work.