附录

图 1 展示了数据集中 3 个样本的图片模态。可以看出，多模态模型准确地描述了设备的外貌、所处环境和功能。由此可以得出，多模态大模型可以生成精良的文本标注，极大减少了人工标注的成本。3 个子图的文本标注如下：

图 1(a)：The image shows two blue mechanical components,..., these components are typically used in piping systems to control the flow of fluids.

图 1(b)：Underneath the vehicle, there is,..., this is a tool commonly used for changing tires or performing maintenance on the vehicle's undercarriage.

图 1(c)：The image shows a large mechanical component that appears to be part of some type of industrial rotary equipment or machine... see a large shaft with several round discs mounted on it. These discs...may be used to transfer power, reduce speed fluctuations, or store energy.

(a) 零部件   (b) 动力工具和手持设备   (c) 重型机械和设备

图 1   数据集部分样本展示

Fig. 1   Partial sample presentation of dataset