

# FastVision-AE: A Low-Latency Explainable Autoencoder for Computer Vision-Based Anomaly Detection

## Abstract

Explanations are crucial for machine learning models in industrial use cases, where understanding model decisions is essential for trust, safety, and effective troubleshooting. However, current explainable AI (XAI) techniques often require extensive computational resources and involve time-consuming processes, leading to long inference times. This makes these methods impractical for real-time anomaly detection applications, where rapid decision-making is critical, such as in industrial inspection or autonomous systems.

To address these limitations, an explainability method that reduces computational complexity and provides near real-time insights is proposed. The presented approach integrates lightweight explainability mechanisms directly into the model, allowing it to effectively detect and localize anomalies. The method was evaluated on industrial datasets, including quality control images, and its performance was compared with existing explainability techniques. The experimental results show that FastVision-AE enables anomaly localization while maintaining minimal inference time overhead. This makes the proposed approach highly suitable for applications requiring rapid decision-making and accurate anomaly detection.