Title : Multimodal Asymmetric Convolutional Autoencoder for Vehicle Small Damage Detection

ABSTRACT

Detecting small vehicle damage, such as scratches, dents, and underbody impacts is an increasingly relevant problem in automotive applications. In car-sharing and rental scenarios, reliable detection is essential for accountability and insurance automation, while in safety systems, recognising subtle surface impacts can complement existing sensors like LiDAR or airbags. Despite its importance, most existing automated approaches rely on camera-based inspection. These methods are constrained by lighting, occlusion, and limited applicability for moving vehicles or underbody inspection, leaving a significant gap for more robust, deployable solutions.

To address this, we investigate inertial (IMU) and audio sensors as complementary modalities for small damage detection. IMUs, traditionally used in airbag systems, capture motion and vibration signatures, while audio sensors record high-frequency transients from scratches and impacts. We propose a Multimodal Asymmetric Autoencoder (MAA) framework that fuses spectrogram representations of both signals at the feature level. The architecture is asymmetric: lightweight convolutional encoders are tailored to each modality, and their latent features are integrated via pooling-based fusion.

Our dataset consists of approximately 3,500 training samples from controlled damage events (augmented from ~500 physical recordings) and over 30,000 test samples collected in diverse driving and environmental conditions. The test set is highly imbalanced (≈40:1 non-damage to damage ratio) and contains unseen damage types as well as realistic non-damage events (potholes, door closures, road bumps), making it a challenging benchmark.

We compare the multimodal framework against mono-modal baselines (IMU-only and audio-only) and explore the role of data representation (raw signals vs. spectrograms). Results show that spectrogram-based representations significantly improve separability between damage and non-damage events, capturing both short-lived impacts and longer scraping patterns. Across experiments, the proposed MAA outperformed unimodal baselines, achieving ROC-AUC 0.92 compared to 0.80 (IMU) and 0.84 (audio). Importantly, the model generalises to unseen damage types and transfers effectively to an external robotics anomaly detection dataset, achieving ROC-AUC 0.89 without tuning.

This work demonstrates that lightweight multimodal fusion, combined with spectrogram-based representations, can bridge the gap between academic anomaly detection research and real-world automotive applications, paving the way for more reliable and deployable small damage detection systems.