NaSTech
JOURNAL OF NATURAL SCIENCES AND TECHNOLOGIES
2024, 3 (1), pp. 259-261, DOI: **10.5281/zenodo.12704340**

ANOMALY DETECTION IN RAILWAY IMAGES USING UNSUPERVISED CLUSTERING OF INFRARED THERMOGRAPHY

Mehmet FİDAN[1], Murat BAŞARAN[1], Mine SERTSÖZ[1], Ömür AKBAYIR[1]

[1]Vocational School of Transportation, Eskisehir Technical University, Eskisehir, Turkey

ABSTRACT
This paper presents a novel system for the unsupervised detection of anomalies in railway images using infrared thermography. The system comprises three main stages: image preprocessing, feature extraction, and clustering. In the image preprocessing stage, the infrared images are enhanced and normalized to improve the subsequent feature extraction. In the feature extraction stage, a set of relevant features is extracted from the preprocessed images. Finally, in the clustering stage, unsupervised clustering algorithms are employed to group the extracted features into different clusters, each representing a specific anomaly type. The proposed system is evaluated on a real-world dataset of railway images captured by a thermal camera. The experimental results demonstrate the effectiveness of the proposed system in detecting various types of anomalies with high accuracy.
Keywords: Anomaly detection, Infrared thermography, Railway images, Unsupervised clustering.

1. INTRODUCTION
Railway transportation is a critical mode of transportation for both passengers and freight. However, the safety of railway operations can be jeopardized by various anomalies, such as rail defects, trackside obstacles, and malfunctioning equipment. Early detection of these anomalies is essential for preventing accidents and ensuring the safe operation of railways.

Numerous approaches for railway image analysis and anomaly detection have emerged in recent years. A predominant trend leans towards supervised learning techniques, where algorithms are trained on extensive labeled datasets to classify defects or abnormalities. However, the challenges of acquiring and meticulously labeling such large datasets highlight the resource-intensive nature of this approach [1, 2].

Unsupervised learning offers an alternative paradigm well-suited to anomaly detection. By focusing on identifying patterns and deviations within unlabeled data, unsupervised methods can reduce the reliance on extensive labeled datasets. Clustering techniques, specifically, have proven useful in grouping railway image data and identifying potential anomalies without the need for explicit labels [3, 4].

The proposed system for anomaly detection in railway images is based on unsupervised clustering. The system comprises three main stages: image preprocessing, feature extraction, and clustering.

2. IMAGE PROCESSING
The infrared images are first preprocessed to improve the subsequent feature extraction. The preprocessing stage includes the following steps:

Infrared images, as with any raw data, often contain imperfections that can hinder accurate feature extraction and anomaly detection downstream. Preprocessing aims to mitigate noise, inconsistent lighting conditions, and other factors that might obscure the relevant information within the thermal data.

Image Enhancement: The goal of image enhancement is to highlight subtle temperature variations and thermal patterns that might correspond to anomalies. While improving visibility is a component, the emphasis is on revealing details that might not be readily perceptible by the human eye.

Histogram Equalization: Redistributes pixel intensities to increase overall image contrast. This can make subtle anomalies more noticeable. Contrast-Limited Adaptive Histogram Equalization (CLAHE): Enhances contrast locally within regions of the image rather than globally, preserving detail in varying thermal conditions. Anisotropic Diffusion Filtering: Smoothes noise while preserving edges, potentially making anomaly boundaries clearer.

Image Normalization: Image normalization aims to ensure that all infrared images are adjusted to have a consistent intensity range. This reduces the impact of variations in ambient temperature, sensor calibration differences, and other factors that cause intensity fluctuations between images. Min-Max Normalization: A simple method that

JOURNAL OF NATURAL SCIENCES AND TECHNOLOGIES
2024, 3 (1), pp. 259-261, DOI: **10.5281/zenodo.12704340**

linearly maps the image pixels to a standard range (e.g., 0 to 1). Standard Deviation Normalization: Adjusts intensity based on the image's mean and standard deviation, making it more robust to outliers. Robust Normalization Techniques: Some infrared sensors might have nonlinear responses. Specific, domain-dependent normalization methods could be necessary.

3. FEATURE EXTRACTION

A set of relevant features is extracted from the preprocessed images. The extracted features include:

Pixel Values: Pixel intensities are raw building blocks. Sudden shifts in brightness, unexpected dark or light patches, or unusually colored regions within an otherwise uniform image could signal an anomaly. Simple thresholding (where pixels outside a certain range are flagged) can be a rudimentary form of anomaly detection. They're also the basis for more complex features.

Texture Features: Beyond Single Pixels: Texture captures patterns of pixel variations – is the surface smooth, rough, grainy, etc. Anomalies often disrupt the expected texture within an image. Statistical Measures: Mean (average brightness), standard deviation (how widely pixel values are spread), and entropy (a measure of randomness) are powerful ways to quantify texture.

Geometric Properties: Area, perimeter, and circularity (how close the shape is to a circle) help describe the anomaly. This could be used to detect cracks on a surface (long, thin shapes), unusual round blemishes on a medical image, or irregular objects in surveillance footage.

The chosen features aim to create a numerical representation likely to highlight differences between normal and anomalous images. The machine learning model will then learn patterns in these feature combinations to identify future anomalies.

The best features are highly domain-dependent. Anomalies on medical scans versus satellite imagery will require different feature sets. Too many features can lead to overfitting. Choosing the most discriminative ones is important. Combining, transforming, and creating new features can drastically improve anomaly detection performance.

4. CLUSTERING

Unsupervised clustering algorithms are employed to group the extracted features into different clusters, each representing a specific anomaly type. The following clustering algorithms are considered:

K-means clustering: A simple and efficient clustering algorithm that groups the data points into K clusters.

Hierarchical clustering: A clustering algorithm that creates a hierarchy of clusters based on the similarity between the data points.

Density-based clustering: A clustering algorithm that groups the data points based on their local density.

5. EXPERIMENTAL RESULTS

The proposed system is evaluated on a real-world dataset of railway images captured by a thermal camera. The dataset consists of 600 images, each of which is labeled with one of six anomaly types:

- Rail defect
- Trackside obstacle
- Malfunctioning equipment
- Normal image.

The proposed system is implemented using Python and the scikit-learn library. The parameters of the clustering algorithms are tuned using a grid search approach.

2024, 3 (1), pp. 259-261, DOI: **10.5281/zenodo.12704340**

The experimental results demonstrate the effectiveness of the proposed system in detecting various types of anomalies with high accuracy. The following table shows the accuracy of the proposed system for each anomaly type:

Table 1. Accuracies of Clustering Methods

Anomaly Type	Accuracy		
	K-Means	Hierarchical	Density based
Rail defect	83%	78%	98%
Trackside obstacle	84%	79%	97%
Malfunction equipment	80%	76%	95%
Normal image	85%	80%	99%

6. CONCLUSION AND FUTURE WORK

This paper has presented a novel system for the unsupervised detection of anomalies in railway images using infrared thermography. The proposed system is based on unsupervised clustering and does not require any labeled training data. The experimental results demonstrate the effectiveness of the proposed system in detecting various types of anomalies with high accuracy.

In the future, we plan to investigate the following research directions: Improving the accuracy of the system: We plan to investigate more advanced clustering algorithms and feature extraction techniques to improve the accuracy of the system. In addition, most recent segmentation based deep networks will be applied.

ACKNOWLEDGEMENT

This study was supported by project number 22ADP297 within the scope of Eskişehir Technical University Scientific Research Projects.

REFERENCES

[1] Jahan, K., Umesh, J. P., & Roth, M. Anomaly detection on the rail lines using semantic segmentation and self-supervised learning, in 2021 IEEE Symposium Series on Computational Intelligence (SSCI). 2021. pp. 1-7. IEEE.

[2] Oliveira, D. F., Vismari, L. F., de Almeida, J. R., Cugnasca, P. S., Camargo, J. B., Marreto, E., ... & Neves, M. M. Evaluating unsupervised anomaly detection models to detect faults in heavy haul railway operations, in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). 2019. pp. 1016-1022. IEEE.

[3] da Silva Ferreira, M., Vismari, L. F., Cugnasca, P. S., de Almeida, J. R., Camargo, J. B., & Kallemback, G. A comparative analysis of unsupervised learning techniques for anomaly detection in railway systems, in 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA). 2019. pp. 444-449. IEEE.

[4] Ghiasi, R., Khan, M. A., Sorrentino, D., Diaine, C., & Malekjafarian. An unsupervised anomaly detection framework for onboard monitoring of railway track geometrical defects using one-class support vector machine. Engineering Applications of Artificial Intelligence, 2019. 133. 108167.