

From Tensors to Novelties: Low-Dimensional Representations for Anomaly Detection in Multispectral Imagery

Anthony C. Chan^{1*}

^{1*}H. Milton Stewart School of Industrial & Systems Engineering
(OMSA – Practicum Summer 2024), Georgia Institute of Technology,
Atlanta, GA, USA.

Corresponding author(s). E-mail(s): anthonycchan@gmail.com;

Abstract

Anomaly detection in image datasets often benefits from reducing high-dimensional inputs into compact, interpretable representations. Tensor decomposition provides a principled way to achieve this by factorizing complex data into low-dimensional components that preserve both spatial and spectral structure. In this work, I investigate how tensor decomposition can be combined with standard anomaly detection models to identify novel patterns in image data. I evaluate two decomposition strategies – global CP (PARAFAC) and per-sample Tucker decompositions – as feature extraction steps. The resulting low-dimensional coefficients are then used to train one-class classifiers, including one-class support vector machines (OC-SVMs), autoencoders (AE), and isolation forests (IF). Models are trained on typical examples only, identifying anomalies as observations poorly explained by the learned representation during training. To demonstrate the methodology, I apply it to multispectral images from the Mars Science Laboratory Curiosity rover, a dataset characterized by scarce labeled anomalies. The experiments reveal a trade-off: CP decomposition provides substantial improvements under locally coherent training conditions, while Tucker decomposition delivers more consistent gains under randomized sampling. These results highlight tensor decomposition as an effective methodology for coupling dimensionality reduction with anomaly detection, with CP suited to structured regimes and Tucker preferred in more heterogeneous settings, with potential applications in planetary exploration, medical imaging, and industrial inspection.

Keywords: tensor decomposition, CP, Tucker, anomaly detection, multispectral data

1 Introduction

Tensor decomposition, such as PARAFAC/CP (Harshman 1970) and Tucker (Tucker 1966), has been used successfully for exploratory multi-way factor analysis in psychometrics and social sciences and continues to be a central tool across disciplines (Kolda and Bader 2009). Anomaly detection aims to identify patterns in data that have not been previously observed (Markou and Singh 2003a,b; Chandola et al. 2009; Pimentel et al. 2014). When combined, these methods can be used to identify novel features in low-dimensional images, such as those collected during rover-based planetary exploration missions. For example, Kerner et al. (2020) compared multiple anomaly detection techniques on multispectral images from the Mars Science Laboratory (MSL) Curiosity rover, demonstrating the importance of unsupervised approaches for prioritizing scientifically interesting observations under constraints of bandwidth and mission time.

Beyond planetary exploration, low-dimensional anomaly detection has been applied in diverse fields, such as the detection of lesions in magnetic resonance imaging (Kim et al. 2021) and the identification of surface cracks in industrial quality-control settings (Tabernik et al. 2019). These examples highlight the versatility of combining dimensionality reduction with anomaly detection in domains where anomalies are rare, labels are scarce, and interpretability is critical.

The present study builds on Kerner et al. (2020) by exploring tensor decomposition as a feature extraction step for anomaly detection in rover-based multispectral images. A key challenge in this domain is the scarcity of labeled novel data, which reflects real-world operational conditions. To address this, I adopt a one-class learning framework: models are trained on typical (non-novel) examples, and novel observations are identified as those poorly explained by the learned representation (Pimentel et al. 2014).

I compare multiple tensor decomposition strategies – global CP (PARAFAC) and per-tile Tucker decomposition – and use the resulting low-dimensional representations as inputs to established anomaly detection models, including one-class support vector machines (SVM), autoencoders, and isolation forests. Performance is evaluated in terms of ROC-AUC under two dataset configurations: an in-order split using the first 1,500 samples for training and a randomized selection drawn from the full dataset. The results show a trade-off between potential and robustness. CP decomposition provides substantial improvements over baselines when the training data is locally coherent, while Tucker decomposition yields more consistent performance across heterogeneous, randomized sampling. These findings demonstrate that tensor decomposition can meaningfully enhance anomaly detection pipelines, with CP suited to structured subsets and Tucker preferred for diverse data regimes.

2 Related Work

2.1 Tensor decompositions

CP/PARAFAC/CANDECOMP. Harshman (1970) introduced the PARAFAC/CP model, which represents a tensor as a sum of rank-1 outer products

with one loading vector per mode, enabling extraction of shared components across modes. **Tucker.** Tucker (1966) proposed the Tucker model, which reduces a tensor to a smaller core coupled with one factor matrix per mode; in this study, the (vectorized) core serves as a compact representation of each training image. **Algorithms and surveys.** An efficient way to compute Tucker factors is via the Higher-Order Singular Value Decomposition (HOSVD), which computes per-mode left singular vectors to assemble orthonormal factor matrices and obtain a core tensor (De Lathauwer et al. 2000). For comprehensive coverage of CP and Tucker notation, algorithms, uniqueness, constraints, and complexity, see Kolda and Bader (2009).

2.2 Anomaly (novelty) detection methods

Surveys and motivation. Chandola et al. (2009); Pimentel et al. (2014) review anomaly/novelty detection, including anomaly taxonomies, application constraints, and evaluation (e.g., ROC-AUC), motivating one-class learning on typical-only data and the use of dimensionality reduction to mitigate high dimensionality. **Canonical methods.** One-class approaches include the One-Class SVM (Schölkopf et al. 2001) and Support Vector Data Description (Tax and Duin 2004); Isolation Forest provides a scalable alternative based on random partitioning (Liu et al. 2008, 2012). In multispectral/hyperspectral imaging, the Reed–Xiaoli (RX) detector is a long-standing baseline (Reed and Yu 1990). **Autoencoders/VAEs and applications.** Unsupervised autoencoder-based detection is common in medical imaging (e.g., brain MRI; (Baur et al. 2021)) and industrial inspection, where the MVTec AD dataset catalyzed benchmarking and emphasized reconstruction- or memory-based approaches (Bergmann et al. 2019, 2021). These works motivate our emphasis on interpretable reconstruction/error maps alongside detector scores.

2.3 Planetary imaging and inspiration for this study

For rover operations, novelty detection supports triage and science targeting under bandwidth and time constraints. Kerner et al. (2020) compared RX, PCA, autoencoders, and GANs on Curiosity Mastcam multispectral data, highlighting performance trade-offs, sensitivities to morphological vs. spectral novelties, and the operational value of explanatory visualizations. The present work is directly inspired by that study and extends it by replacing end-to-end pixel representations with tensor-derived features (CP and Tucker) fed to one-class detectors; it also uses the datasets provided by Kerner et al. (2020).

3 Methodology

3.1 Dataset Partitioning

The Kerner et al. (2020) dataset was downsampled to create two smaller subsets for faster training. The full dataset consists of 9,302 training tiles, 1,386 validation tiles, and a test set with 426 typical (non-anomalous) tiles and 430 novel (anomalous) tiles.

The first subset uses the first 1,500 training tiles, the first 200 validation tiles, and the first 200 typical plus the first 200 novel test tiles. Ordering is based on the

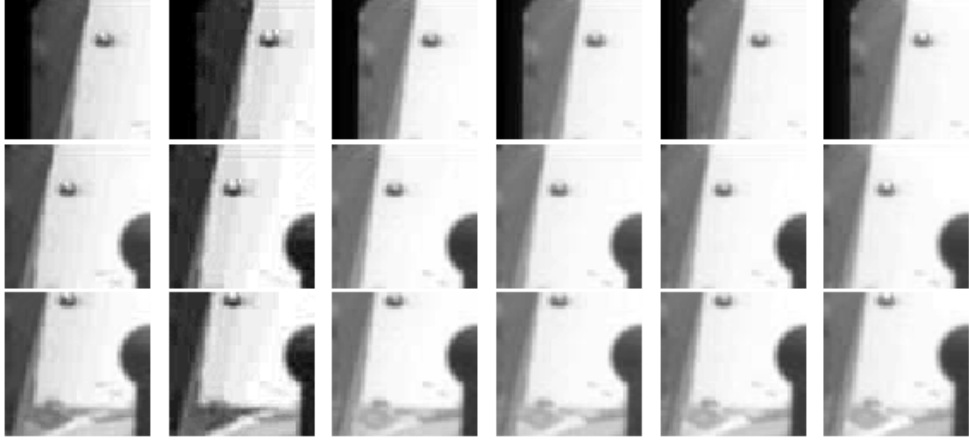


Fig. 1 Sample tiles from the original dataset: typical (non-anomalous).

alphanumeric sequence of file names, which also corresponds to the chronological order of data collection.

The second subset is constructed using a randomized sampling of 1,500 training tiles, 200 validation tiles, 200 typical test tiles, and 200 novel test tiles from the original dataset.

3.2 Global CP (PARAFAC) Decomposition

I fit a global CANDECOMP/PARAFAC (CP) model *only on the training tensor*. Let the training stack be a fourth-order tensor $\mathcal{X}_{\text{train}} \in \mathbb{R}^{N \times I \times J \times K}$ with $N = 1500$, $I = J = 64$, $K = 6$. The model approximates

$$\mathcal{X}_{\text{train}} \approx \sum_{r=1}^R \lambda_r h_r \circ a_r \circ b_r \circ c_r,$$

yielding factor matrices $H \in \mathbb{R}^{N \times R}$, $A \in \mathbb{R}^{I \times R}$, $B \in \mathbb{R}^{J \times R}$, $C \in \mathbb{R}^{K \times R}$, and weights $\lambda \in \mathbb{R}^R$. Columns of A, B, C capture shared spatial and spectral patterns, while each row of H contains per-tile coefficients. The compression level (rank R) is chosen during anomaly detection model training via a grid search with the practical constraint $R \leq \min(IJ, JK, IK) = 385$ (see, e.g., (Kolda and Bader 2009; Bro 1997)). Validation and test tiles are *not refit*; they are projected into the training CP space defined by (A, B, C) .

To obtain scale-consistent per-sample coefficients, we absorb the component weights λ into the sample-mode factor:

$$\tilde{H} = H \text{diag}(\lambda),$$

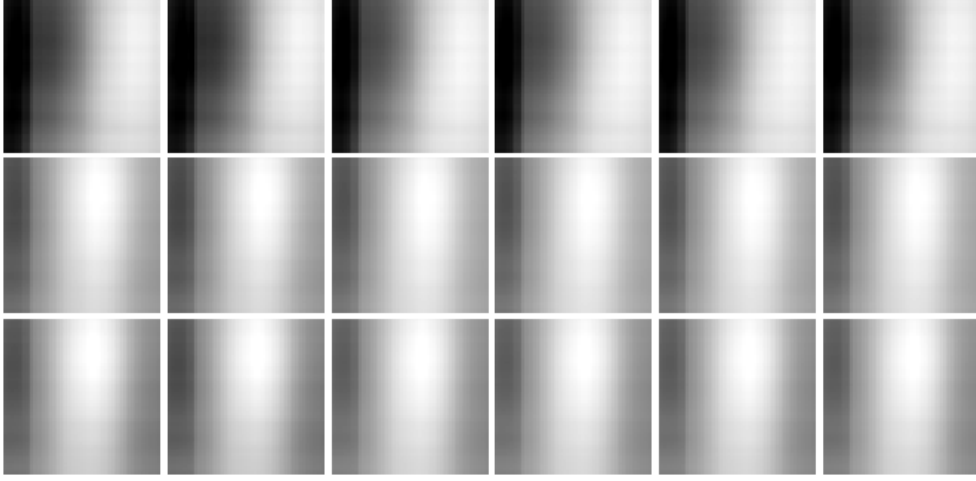


Fig. 2 CP reconstructions at rank $R = 5$.

so that each row of \tilde{H} reflects the absolute contribution of each component across samples (without imposing nonnegativity constraints).

For a *single* tile $X \in \mathbb{R}^{I \times J \times K}$ from validation or test, define component overlaps

$$g_r(X) = \langle X, a_r \circ b_r \circ c_r \rangle, \quad r = 1, \dots, R,$$

and the Hadamard Gram matrix

$$G = (A^\top A) * (B^\top B) * (C^\top C),$$

where \circ denotes the outer product and $*$ the elementwise product. We obtain CP coefficients by solving a regularized normal equation

$$z(X) = (G + \epsilon I)^{-1} g(X) \in \mathbb{R}^R,$$

with $\epsilon = 10^{-6}$ to mitigate ill-conditioning at high ranks. Stacking rows $z(X_n)^\top$ over a batch yields $Z \in \mathbb{R}^{N \times R}$ (Kiers 2000; Sørensen and Bro 2009).

3.3 Tucker Decomposition

Tucker decomposition is applied to the training, validation and test datasets for each image tile $X_n \in \mathbb{R}^{64 \times 64 \times 6}$ given the ranks (r_1, r_2, r_3) . The Tucker model represents each tile via a core tensor and three orthonormal bases (one per mode) (Tucker 1966; Kolda and Bader 2009):

$$X_n \approx \hat{X}_n = \mathcal{G}_n \times_1 U_1 \times_2 U_2 \times_3 U_3, \quad U_i^\top U_i = I_{r_i},$$

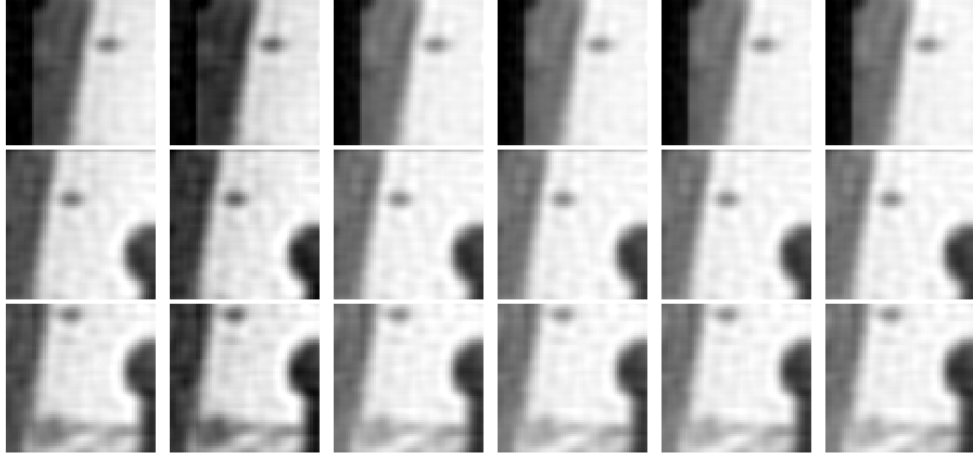


Fig. 3 CP reconstructions at rank $R = 200$.

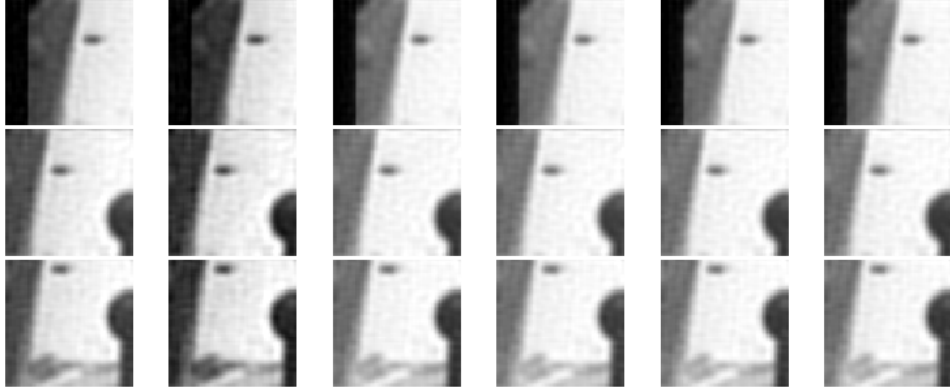


Fig. 4 CP reconstructions at rank $R = 380$.

with dimensions

$$\mathcal{G}_n \in \mathbb{R}^{r_1 \times r_2 \times r_3}, \quad U_1 \in \mathbb{R}^{64 \times r_1}, \quad U_2 \in \mathbb{R}^{64 \times r_2}, \quad U_3 \in \mathbb{R}^{6 \times r_3}.$$

Here, \times_n denotes the mode- n tensor-matrix product.

To perform the decomposition of the image tiles, form a tensor \mathcal{X} by stacking the image tiles along a sample mode and compute the mode- n unfoldings $\mathcal{X}_{(n)}$. The orthonormal bases are obtained from the left singular vectors of these unfoldings (truncated to rank

r_n) via the Higher-Order SVD (HOSVD) (De Lathauwer et al. 2000):

$$\begin{aligned}\mathcal{X}_{(1)} &= U_1 S_1 V_1^\top \Rightarrow U_1 = U_1(:, 1:r_1), \\ \mathcal{X}_{(2)} &= U_2 S_2 V_2^\top \Rightarrow U_2 = U_2(:, 1:r_2), \\ \mathcal{X}_{(3)} &= U_3 S_3 V_3^\top \Rightarrow U_3 = U_3(:, 1:r_3).\end{aligned}$$

With U_1, U_2, U_3 fixed, each tile is projected to its core by three orthogonal projections (Kolda and Bader 2009):

$$\boxed{\mathcal{G}_n = X_n \times_1 U_1^\top \times_2 U_2^\top \times_3 U_3^\top}.$$

This yields a compressed representation that retains dominant multilinear structure while attenuating small, noisy variation (Kolda and Bader 2009). Each core is vectorized to produce a feature vector

$$z_n = \text{vec}(\mathcal{G}_n) \in \mathbb{R}^{r_1 r_2 r_3},$$

and, for a batch of N tiles, the feature matrix is

$$Z = \begin{bmatrix} z_1^\top \\ \vdots \\ z_N^\top \end{bmatrix} \in \mathbb{R}^{N \times (r_1 r_2 r_3)}.$$

These vectors z_n serve as inputs to the downstream anomaly detection models.

Rank is chosen based on the dimension of the tensor being decomposed. In this case, the image dataset has a dimension of $64 \times 64 \times 6$. To use the full spatial dimensionality of each factor, a search up to 64 captures the spatial variation of the first two modes; for the third (spectral) mode, up to rank 6 captures its full variation.

3.4 Standardization

Given the decomposed training, validation, and test datasets for both CP and Tucker, I standardize the feature vectors and use the standardized features for downstream anomaly detection (one-class SVM, autoencoder, and Isolation Forest) to predict whether an image tile is typical or novel. The per-feature z-score transform is

$$\tilde{z} = S(z) = \frac{z - \mu_{\text{train}}}{\sigma_{\text{train}}} \in \mathbb{R}^d,$$

where $z \in \mathbb{R}^d$ is the feature vector for an image tile, and $\mu_{\text{train}} \in \mathbb{R}^d$, $\sigma_{\text{train}} \in \mathbb{R}^d$ are the mean and standard deviation computed on the *training* features. The same transformation $S(\cdot)$ is applied to validation and test features (without refitting) to avoid information leakage.

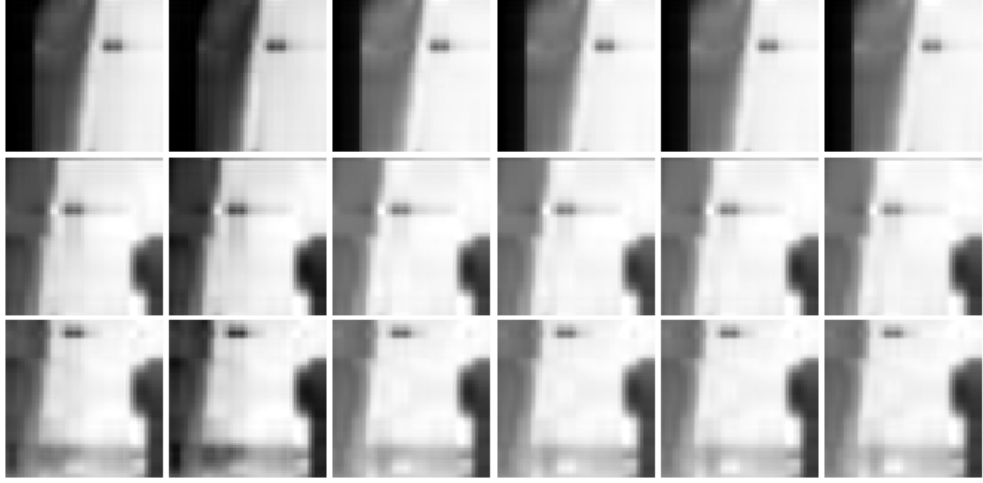


Fig. 5 Tucker reconstructions for $(5, 5, 5)$.

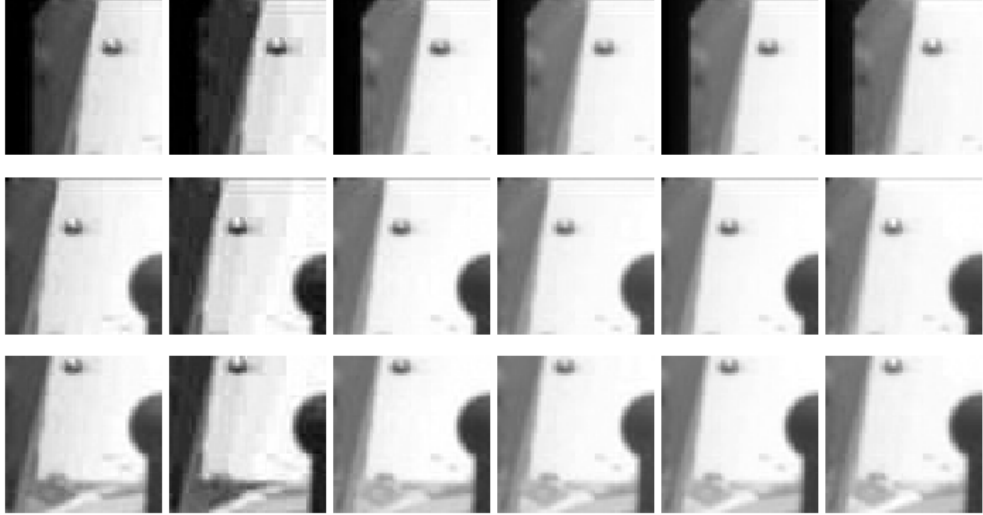


Fig. 6 Tucker reconstructions for $(32, 32, 5)$.

3.5 One-Class SVM (RBF Kernel)

Let $\tilde{Z}_{\text{train}} \in \mathbb{R}^{|T| \times d}$ denote the standardized training feature matrix (rows are tiles; columns are CP/Tucker features). I train an OC-SVM model with an RBF kernel $K(z, z') = \exp\{-\gamma\|z - z'\|_2^2\}$ (Schölkopf et al. 2001; Tax and Duin 2004), where it

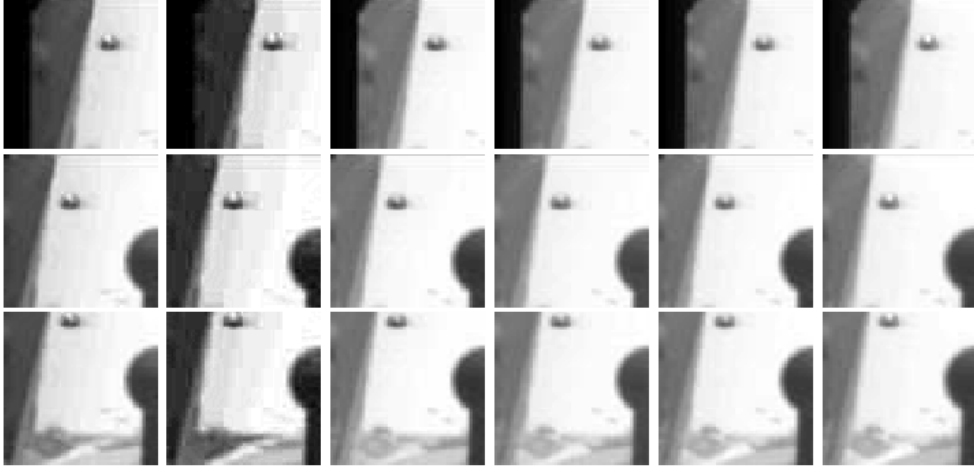


Fig. 7 Tucker reconstructions for (64, 64, 16).

solves the OC-SVM dual formulation

$$\max_{\alpha} -\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j K(z_i, z_j) \quad \text{s.t.} \quad 0 \leq \alpha_i \leq \frac{1}{\nu n}, \quad \sum_{i=1}^n \alpha_i = 1.$$

In the above model, $z_i \in \mathbb{R}^d$ denotes a training feature vector (rows of the training matrix $Z \in \mathbb{R}^{n \times d}$), and ν is a hyperparameter.

Since validation data is typical-only, I minimize the false-positive rate at the default OC-SVM decision boundary $f(z) = 0$ (Schölkopf et al. 2001; Tax and Duin 2004). Following common practice in one-class/unsupervised evaluation (Chandola et al. 2009; Campos et al. 2016; Ruff et al. 2021), I minimize the average indicator of nonnegative validation scores:

$$\min \frac{1}{|V|} \sum_{z \in V} \mathbf{1}[s_{\text{val}}(z) \geq 0],$$

where V is the validation dataset and $s_{\text{val}}(z)$ is the OC-SVM validation score for sample z , used for anomaly decisions:

$$s_{\text{val}}(z) = \sum_{i=1}^n \alpha_i K(z_i, z) - \rho.$$

The model is tuned over the parameter grid

$$\gamma \in \left\{ \frac{t}{d} : t \in \{0.1, 0.3, 1, 3, 10\} \right\} \cup \{\text{scale}, \text{auto}\}, \quad \nu \in \{0.01, 0.02, 0.05, 0.10, 0.20\},$$

and evaluated using the validation scores. The rank/hyperparameters are selected as described below; the selected setting is then applied to the test dataset to compute

ROC-AUC, and I report the maximum accuracy achievable under that operating point.

3.6 Autoencoders

Let $\tilde{Z}_{\text{train}} \in \mathbb{R}^{|T| \times d}$ denote the standardized training feature matrix (rows are tiles; columns are CP/Tucker features). I train a four-layer, fully connected autoencoder with hidden layer size $H \in \{128, 256, 384\}$ and bottleneck size $B \in \{16, 32, 64\}$: the encoder maps the feature dimension $\mathbb{R}^d \rightarrow \mathbb{R}^H \rightarrow \mathbb{R}^B$ and the decoder mirrors this back to $\mathbb{R}^H \rightarrow \mathbb{R}^d$ (Hinton and Salakhutdinov 2006; Goodfellow et al. 2016). Encoder is represented as follows:

$$h_{\text{enc}} = E_{\theta}(\tilde{z}) \in \mathbb{R}^B.$$

Decoder is represented as follows:

$$\hat{z} = D_{\theta}(h_{\text{enc}}) \in \mathbb{R}^d.$$

Model parameters θ are learned by minimizing mean squared reconstruction error on \tilde{Z}_{train} , a standard objective for autoencoders (Hinton and Salakhutdinov 2006; Goodfellow et al. 2016):

$$\theta^* = \arg \min_{\theta} \frac{1}{|T|} \sum_{\tilde{z} \in Z_{\text{train}}} \frac{1}{d} \|\tilde{z} - f_{\theta}(\tilde{z})\|_2^2.$$

The anomaly score for a sample \tilde{z} is its normalized reconstruction error:

$$s(\tilde{z}) = \frac{1}{d} \|\tilde{z} - f_{\theta}(\tilde{z})\|_2^2,$$

Larger values indicate greater deviation from normal reconstructions (Sakurada and Yairi 2014; Zhou and Paffenroth 2017). Since the validation set contains only typical samples, the operating threshold is set by a high validation quantile,

$$\tau_{\text{val}} = \text{P95}(\{s(\tilde{z}) : \tilde{z} \in Z_{\text{val}}\}),$$

and predictions follow

$$\hat{y}(x) = \mathbf{1}[s(x) \geq \tau_{\text{val}}],$$

a common calibration strategy for one-class/unsupervised anomaly detection (Chandola et al. 2009; Campos et al. 2016; Ruff et al. 2021). *Rank/hyperparameter selection:* the decomposition rank is chosen on the typical-only validation set by lexicographic minimization of FP@0, then P95, then the mean of $s(\tilde{z})$; τ_{val} is fixed for test evaluation.

3.7 Isolation Forest

Isolation Forest builds an ensemble of binary trees on subsamples of the training feature vectors and exploits the fact that anomalies are few and different: they are

isolated by random splits in fewer steps than typical points (Liu et al. 2008, 2012). Each tree is grown by recursively selecting a feature and a split value uniformly at random and partitioning the node until a single point remains or a depth limit is reached (Liu et al. 2008).

We explore the following hyperparameter grid:

$$T \in \{50, 100, 200\}, \quad p_{\text{sub}} \in \{0.5, 0.75, 1.0\}, \quad m_f \in \{0.5, 0.75, 1.0\}, \quad \text{contamination} \in \{0.05, 0.10, 0.20\},$$

where T is the number of trees, p_{sub} is the subsample fraction per tree, m_f is the fraction of features considered per split, and the contamination parameter specifies an assumed anomaly fraction for the model’s default decision threshold (relevant only when using the built-in `predict`) (Liu et al. 2012). Let $n_{\text{sub}} = \lfloor p_{\text{sub}} \cdot |T| \rfloor$ denote the resulting subsample size per tree.

For a feature vector z , the anomaly score is

$$s(z) = 2^{-\bar{h}(z)/c(n_{\text{sub}})}, \quad \bar{h}(z) = \frac{1}{T} \sum_{t=1}^T h_t(z),$$

where $h_t(z)$ is the path length of z in tree t , and $c(n_{\text{sub}})$ normalizes scores across subsample sizes (Liu et al. 2008):

$$c(n_{\text{sub}}) = 2 H_{n_{\text{sub}}-1} - \frac{2(n_{\text{sub}}-1)}{n_{\text{sub}}}, \quad H_m = \sum_{k=1}^m \frac{1}{k} \approx \ln(m) + \gamma,$$

with H_m the m -th harmonic number and γ the Euler–Mascheroni constant.

Because the validation set contains only typical samples, we calibrate the operating threshold by a high validation quantile (Chandola et al. 2009; Campos et al. 2016; Ruff et al. 2021):

$$\tau_{\text{val}} = \text{P95}(\{s(z_v) : z_v \in V\}), \quad \hat{y}(z) = \mathbf{1}[s(z) \geq \tau_{\text{val}}],$$

so that approximately 5% of validation points are flagged as anomalous. *Rank/hyperparameter selection:* we select by the same lexicographic rule on typical-only validation scores (FP@0 \rightarrow P95 \rightarrow mean), then fix τ_{val} for test evaluation.

4 Experiments

4.1 Overview

In Section 3, I define dataset partitioning, the CP and Tucker decompositions, and feature standardization. The following sections describe the model training and tests performed using the common decomposed training, validation, and test datasets. I consolidate the discussions for models using CP and Tucker decompositions since they share the same anomaly-detection models and differ only in the input decomposition technique. Two dataset configurations are evaluated: (i) an *in-order* split using the

Model	Test set	bedrock	broken-rock	drill-hole	drt	dump-pile
OC-SVM	0.5625	0.2918	0.4070	0.4680	0.6420	0.4940
AE	0.5418	0.5372	0.5971	0.4604	0.4724	0.4512
IF	0.6094	0.8512	0.8743	0.3301	0.2454	0.3163

Table 1 Performance of baseline models across categories up to *dump-pile*.

Model	float	meteorite	scuff	veins
OC-SVM	0.2990	0.4820	0.1320	0.7630
AE	0.6821	0.4689	0.5486	0.6911
IF	0.8889	0.6894	0.6042	0.9300

Table 2 Performance of baseline models across categories from *float* to *veins*.

first 1,500 training tiles, and (ii) a *randomized* selection drawn from the full dataset to assess robustness to sampling diversity.

4.2 CP/Tucker Rank Search

For CP decomposition, I sweep the rank $R \in \{5, 10, 15, \dots, 380\}$. For Tucker decomposition, I sweep (R_1, R_2, R_3) where $R_1, R_2 \in \{5, 16, 32, 64\}$ and $R_3 \in \{5, 16\}$. The decomposition rank (CP: R ; Tucker: (R_1, R_2, R_3)) is selected *exclusively* on the typical-only validation set using a lexicographic criterion applied to validation anomaly scores: minimize FP@0, then P95, then the mean. Test labels are used only for final reporting (ROC-AUC, accuracy). The optimal ranks for each setting are summarized in Tables 3 and 4 in Section 5.

5 Results

This study evaluated three standard anomaly detection models – OC-SVM, autoencoders (AE), and Isolation Forest (IF). These models were trained both on raw pixel data and on data decomposed using tensor methods. Performance was assessed under two setups as described in Section 3: (i) an in-order split using the first 1,500 datapoints for training and validation, and (ii) a randomized selection from the full dataset.

5.1 Baseline Models

OC-SVM, autoencoders, and Isolation Forest models were trained and validated on the reduced dataset defined in Section 3. These baseline models, used without tensor decomposition, provide a reference for the composite models discussed later. Tables 1 and 2 present ROC-AUC values for the baseline models evaluated on the test set and the anomaly categories.

Model	Best Rank(s)
CP + OC-SVM	120
Tucker + OC-SVM	(32, 32, 16)
CP + AE	35
Tucker + AE	(64, 32, 5)
CP + IF	365
Tucker + IF	(64, 16, 5)

Table 3 Optimal ranks (in-order split).

Model	Best Rank(s)
CP + OC-SVM	15
Tucker + OC-SVM	(64, 16, 5)
CP + AE	230
Tucker + AE	(64, 64, 5)
CP + IF	315
Tucker + IF	(64, 5, 5)

Table 4 Optimal ranks (randomized split).

5.2 Rank Selection Summary

Tables 3 and 4 summarize the optimal ranks selected by validation for each decomposition–detector combination under the in-order and randomized configurations. All results reported below use these selected ranks.

5.3 CP Decomposition

The CP results reported below use the optimal ranks in Tables 3 and 4.

For the in-order split (Tables 5 and 6), CP + OC-SVM improved ROC–AUC from 0.56 to 0.71 on the test set, with substantial gains in categories such as bedrock, broken-rock, float, meteorite, scuff, and veins. CP + AE showed more modest gains on the test set (0.58 vs. 0.54 baseline), with improvements across several categories but declines for bedrock, float, and veins. CP + IF improved slightly on the test set (0.64 vs. 0.60 baseline), though performance declined for bedrock, broken-rock, float, meteorite, and veins categories.

For the randomized selection (Tables 7 and 8), CP decomposition generally underperformed compared to the raw-pixel baselines. CP + OC-SVM dropped from 0.56 to 0.49, CP + AE from 0.64 to 0.54, and CP + IF from 0.65 to 0.50, with only minor gains observed in isolated categories such as broken-rock.

5.4 Tucker Decomposition

The Tucker results reported below use the optimal ranks in Tables 3 and 4. Tucker decomposition produced modest improvements under the in-order split but outperformed CP under randomized selection.

Model	Test set	bedrock	broken-rock	drill-hole	drt	dump-pile
CP + OC-SVM	0.7095	0.6860	0.8002	0.5625	0.3681	0.4768
CP + AE	0.5778	0.5321	0.6055	0.6597	0.6389	0.5662
CP + IF	0.6487	0.7567	0.9256	0.1597	0.1944	0.3311

Table 5 Performance of in-order CP-based models across categories up to *dump-pile*.

Model	float	meteorite	scuff	veins
CP + OC-SVM	0.6697	0.6817	0.4722	0.9156
CP + AE	0.4969	0.7076	0.6597	0.6311
CP + IF	0.9414	0.7803	0.4097	0.9411

Table 6 Performance of in-order CP-based models across categories from *float* to *veins*.

Model	Test set	bedrock	broken-rock	drill-hole	drt	dump-pile
CP + OC-SVM	0.4914	0.4136	0.4951	0.5023	0.5569	0.4762
CP + AE	0.5405	0.6514	0.7751	0.4711	0.3114	0.3495
CP + IF	0.5048	0.6878	0.7935	0.4029	0.3033	0.3216

Table 7 Performance of randomized CP-based models across categories up to *dump-pile*.

For the in-order split (Tables 9 and 10), Tucker + OC-SVM showed a small improvement from 0.54 to 0.57 on the test set, with gains in bedrock, broken-rock, float, meteorite, and scuff. Tucker + AE improved from 0.54 to 0.59, with gains across most categories except drt. Tucker + IF slightly declined from 0.61 to 0.60 on the test set, with performance losses in bedrock, broken-rock, float, meteorite, and veins.

For the randomized selection (Tables 11 and 12), Tucker + OC-SVM improved from 0.56 to 0.64, with gains in bedrock, broken-rock, float, meteorite, scuff, and veins. Tucker + AE increased from 0.64 to 0.66 on the test set, though drill-hole, drt, dump-pile, and meteorite declined. Tucker + IF dropped sharply from 0.65 to 0.49, with performance losses across most categories.

6 Discussion

The experiments highlight both the potential and the limitations of tensor decompositions as feature extraction strategies for anomaly detection in multispectral imagery. Two evaluation setups were considered: (i) an in-order split using the first 1,500 datapoints for training, and (ii) a randomized selection drawn from the full dataset. Together, these provide insight into the comparative strengths of CP and Tucker decompositions.

Model	float	meteorite	scuff	veins
CP + OC-SVM	0.4356	0.5218	0.5042	0.5500
CP + AE	0.7292	0.6393	0.4700	0.8088
CP + IF	0.7606	0.6235	0.4458	0.8233

Table 8 Performance of randomized CP-based models across categories from *float* to *veins*.

Model	Test set	bedrock	broken-rock	drill-hole	drt	dump-pile
Tucker + OC-SVM	0.5710	0.7025	0.8131	0.3889	0.3264	0.3842
Tucker + AE	0.5900	0.6979	0.7898	0.4653	0.4514	0.4589
Tucker + IF	0.6028	0.7968	0.7569	0.5556	0.5000	0.4678

Table 9 Performance of Tucker-based models across categories up to *dump-pile*.

Model	float	meteorite	scuff	veins
Tucker + OC-SVM	0.7346	0.5363	0.6250	0.7567
Tucker + AE	0.7160	0.5052	0.6597	0.7956
Tucker + IF	0.7469	0.6367	0.7083	0.7689

Table 10 Performance of Tucker-based models across categories from *float* to *veins*.

6.1 Improvements Under In-Order Splits

When trained on the first 1,500 datapoints in sequence, CP decomposition produced substantial improvements over baseline models. CP + OC-SVM in particular raised overall ROC-AUC from 0.56 (raw OC-SVM) to 0.71, with large category-level gains for bedrock, broken-rock, and veins (up to 0.40 AUC improvement). These results suggest that CP’s global rank structure is highly effective at capturing dominant spatial-spectral patterns when training data is locally coherent, enhancing separability for downstream classifiers such as OC-SVM and Isolation Forest. The optimal ranks identified for these models ($R = 120$ for CP + OC-SVM, $R = 35$ for CP + AE, and $R = 365$ for CP + IF) suggest that moderate to high global ranks were required to achieve maximal separation, emphasizing the capacity of CP to represent locally consistent spatial-spectral structure.

6.2 Robustness Under Randomized Sampling

In contrast, when training samples were drawn at random across the full dataset, CP decomposition underperformed relative to baselines. CP + OC-SVM dropped to 0.49 ROC-AUC, below raw OC-SVM (0.56), with similar degradations for CP + AE and CP + IF. This indicates that CP’s single-rank factorization, while effective on homogeneous subsets, struggles to generalize under heterogeneous sampling.

Tucker decomposition, however, demonstrated greater stability. Tucker + OC-SVM achieved 0.64 ROC-AUC, surpassing its baseline (0.56), while Tucker + AE

Model	Test set	bedrock	broken-rock	drill-hole	drt	dump-pile
Tucker + OC-SVM	0.6362	0.8705	0.8872	0.5094	0.4193	0.5654
Tucker + AE	0.6643	0.8300	0.9055	0.5486	0.4512	0.5762
Tucker + IF	0.4925	0.5027	0.5403	0.3977	0.4230	0.5076

Table 11 Performance of Tucker-based models across categories up to *dump-pile*.

Model	float	meteorite	scuff	veins
Tucker + OC-SVM	0.7692	0.7128	0.8938	0.8638
Tucker + AE	0.7961	0.6662	0.9100	0.9152
Tucker + IF	0.5075	0.3699	0.6850	0.6362

Table 12 Performance of Tucker-based models across categories from *float* to *veins*.

also improved slightly (0.66 vs. 0.64). Gains were evident across multiple categories, though performance stagnated or declined for drill-hole, drt, and dump-pile. The optimal ranks for Tucker under randomized sampling – particularly (64, 16, 5) for OC-SVM and (64, 64, 5) for AE – indicate that higher spatial ranks improved flexibility while keeping compact spectral dimensions, contributing to its stability under heterogeneous conditions. These results highlight Tucker’s ability to flexibly allocate rank across modes, distributing spectral and spatial variability in a way that generalizes more effectively to diverse data. Isolation Forest, while strong as a baseline, showed limited benefit or degradation when paired with decompositions, suggesting it already captures variance structure effectively in raw feature space.

6.3 Category-Specific Recommendations

The category-level analysis suggests that CP and Tucker decompositions offer complementary strengths. For locally coherent data, CP + OC-SVM yielded substantial improvements in anomalies such as bedrock, broken-rock, and veins, where global rank structure captures shared spectral-spatial patterns. For more heterogeneous anomaly types, Tucker decomposition proved more reliable, consistently outperforming baselines for categories including bedrock, broken-rock, float, scuff, and veins under randomized sampling. These findings indicate that decomposition choice can be tailored to the anomaly type of interest rather than applied as a one-size-fits-all strategy. These category-specific differences also align with the rank-search outcomes: CP required larger ranks to capture localized consistency, while Tucker achieved strong performance with compact yet balanced rank configurations.

6.4 Trade-Off Between Potential and Stability

Overall, the two experiments reveal a trade-off between potential gain and robustness. CP decomposition provides striking improvements when training data is locally consistent and representative of test conditions, but its performance deteriorates under randomized sampling. Tucker decomposition delivers more modest improvements in

structured splits yet maintains stable or superior performance under heterogeneous conditions. From a practical perspective, this robustness makes Tucker decomposition a stronger candidate for deployment in real-world scenarios where training data must capture diverse and unpredictable distributions.

6.5 Considerations

While the current experiments demonstrate the utility of CP and Tucker decompositions for anomaly detection in multispectral imagery, several directions remain open. First, training was performed on relatively limited subsets (e.g., 1,500 data-points). Expanding the training pool could enable both CP and Tucker models to learn more representative latent factors, mitigating CP’s sensitivity to heterogeneity and strengthening generalization.

Second, the number of anomaly test samples was limited. Incorporating additional labeled anomalies, either through annotation or synthetic augmentation, would allow more robust estimation of model sensitivity across categories and clarify whether CP’s advantages on structured anomalies and Tucker’s stability under heterogeneous data persist at scale.

Finally, tensor decomposition methods should be applied to additional datasets. While this study focused on multispectral rover imagery, CP and Tucker factorizations are general techniques that extend to other domains such as medical imaging (MRI, CT) and industrial inspection. Benchmarking across diverse datasets will help determine whether the category-specific advantages observed here generalize and are broadly applicable.

7 Conclusion

This work evaluated the use of tensor decomposition methods, specifically CP and Tucker factorizations, as feature extraction strategies for anomaly detection in multispectral imagery. By comparing models trained on raw pixel data with those trained on decomposed representations, two consistent findings emerged.

First, CP decomposition demonstrated strong potential under locally coherent training conditions. In in-order splits, CP + OC-SVM in particular achieved substantial improvements, raising ROC-AUC by as much as 40% in certain anomaly categories. These results highlight the power of global CP structure in capturing dominant spectral-spatial patterns when training and test distributions are aligned. The optimal ranks identified for the in-order setup ($R = 120$ for CP + OC-SVM and $R = 365$ for CP + IF) further underscore the role of high-rank global structure in modeling consistent data.

Second, Tucker decomposition exhibited greater robustness under heterogeneous conditions. In randomized sampling, Tucker-based models consistently outperformed both baselines and CP-based pipelines, achieving the highest overall test set ROC-AUC and providing reliable gains across multiple categories. These outcomes suggest that Tucker’s ability to flexibly allocate rank across modes makes it more adaptable to diverse data distributions. Across experiments, the optimal rank search revealed that

CP models favored higher ranks (e.g., 120–365), while Tucker models achieved competitive or superior results with lower, mode-specific ranks (e.g., 32, 32, 16 or 64, 16, 5), underscoring Tucker’s efficiency in representing heterogeneous structure.

Taken together, the results indicate a trade-off between potential and stability. CP offers the greatest performance gains in structured regimes, while Tucker provides more reliable performance in settings where variability and heterogeneity are high. From an applied perspective, this suggests that the choice of decomposition strategy should depend on the data characteristics and the anomaly categories of interest.

Future work will involve scaling the training sets, expanding anomaly annotations, and applying these methods to additional domains such as medical imaging or industrial inspection. Such efforts will help determine whether the category-specific strengths observed here generalize across broader contexts.

Overall, this study shows that tensor decompositions can substantially enhance anomaly detection pipelines, provided their use is tailored to the structure of the data and the deployment setting.

Acknowledgments

This work was conducted as part of the author’s practicum project in the Online Master of Science in Analytics (OMSA) program at the Georgia Institute of Technology. The author gratefully acknowledges the Mars multispectral image dataset prepared and analyzed by (Kerner et al. 2020), which enabled this study. The author also thanks the OMSA faculty and practicum advisors for their guidance during the project.

Statements and Declarations

Funding

Not applicable.

Competing Interests

The author declares that no competing interests exist.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Data availability

The experiments used the “Mars novelty detection Mastcam labeled dataset” (Zenodo, DOI: 10.5281/zenodo.3732485). No new datasets were generated.

Code availability

All code and configuration to reproduce the experiments are available at https://github.com/anthonyccchan/ISyE6748_practicum_publish/

Author contribution

The author confirms being the sole contributor of this work and has approved it for publication.

References

- Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C.: The MVTec anomaly detection dataset: A comprehensive real-world dataset for unsupervised anomaly detection **129**, 1038–1059 (2021) <https://doi.org/10.1007/s11263-020-01400-4>
- Baur, C., Denner, S., Wiestler, B., Albarqouni, S., Navab, N.: Autoencoders for unsupervised anomaly segmentation in brain MR images: A comparative study **69**, 101952 (2021) <https://doi.org/10.1016/j.media.2020.101952>
- Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: MVTec AD—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9592–9600 (2019). <https://doi.org/10.1109/CVPR.2019.00983>
- Bro, R.: PARAFAC: Tutorial and applications **38**(2), 149–171 (1997) [https://doi.org/10.1016/S0169-7439\(97\)00032-4](https://doi.org/10.1016/S0169-7439(97)00032-4)
- Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: A survey **41**(3), 15–11558 (2009) <https://doi.org/10.1145/1541880.1541882>
- Campos, G.O., Zimek, A., Sander, J., Campello, R.J.G.B., Micenková, B., Schubert, E., Assent, I., Houle, M.E.: On the evaluation of unsupervised outlier detection: Measures, datasets, and an empirical study **30**(4), 891–927 (2016) <https://doi.org/10.1007/s10618-015-0444-8>
- De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition **21**(4), 1253–1278 (2000) <https://doi.org/10.1137/S0895479896305696>
- Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, ??? (2016). <http://www.deeplearningbook.org>
- Harshman, R.A.: Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis **16**, 1–84 (1970)
- Hinton, G.E., Salakhutdinov, R.R.: Reducing the dimensionality of data with neural networks **313**(5786), 504–507 (2006) <https://doi.org/10.1126/science.1127647>

- Kolda, T.G., Bader, B.W.: Tensor decompositions and applications **51**(3), 455–500 (2009) <https://doi.org/10.1137/07070111X>
- Kim, M., Cho, J., Lee, S.: Unsupervised anomaly detection for brain MR images using deep learning: A survey **86**(2), 627–648 (2021) <https://doi.org/10.1002/mrm.28750>
- Kiers, H.A.L.: Towards a standardized notation and terminology in multiway analysis **14**(3), 105–122 (2000) [https://doi.org/10.1002/1099-128X\(200005/06\)14:3<105::AID-CEM582>3.0.CO;2-I](https://doi.org/10.1002/1099-128X(200005/06)14:3<105::AID-CEM582>3.0.CO;2-I)
- Kerner, H.R., Wagstaff, K.L., Bue, B.D., et al.: Comparison of novelty detection methods for multispectral images in rover-based planetary exploration missions **34**(6), 1837–1862 (2020) <https://doi.org/10.1007/s10618-020-00697-6>
- Kerner, H., Wellington, D., Wagstaff, K., Bell, J., Ben Amor, H.: Mars Novelty Detection Mastcam Labeled Dataset. <https://doi.org/10.5281/zenodo.3732485> . Licensed under CC BY 4.0. <https://doi.org/10.5281/zenodo.3732485>
- Liu, F.T., Ting, K.M., Zhou, Z.-H.: Isolation forest. In: Proceedings of the 2008 IEEE International Conference on Data Mining (ICDM), pp. 413–422 (2008). <https://doi.org/10.1109/ICDM.2008.17>
- Liu, F.T., Ting, K.M., Zhou, Z.-H.: Isolation-based anomaly detection **6**(1), 3–1339 (2012) <https://doi.org/10.1145/2133360.2133363>
- Markou, M., Singh, S.: Novelty detection: A review—part 1: Statistical approaches **83**(12), 2481–2497 (2003) <https://doi.org/10.1016/j.sigpro.2003.07.018>
- Markou, M., Singh, S.: Novelty detection: A review—part 2: Neural network based approaches **83**(12), 2499–2521 (2003) <https://doi.org/10.1016/j.sigpro.2003.07.019>
- Pimentel, M.A.F., Clifton, D.A., Clifton, L., Tarassenko, L.: A review of novelty detection **99**, 215–249 (2014) <https://doi.org/10.1016/j.sigpro.2013.12.026>
- Ruff, L., Kauffmann, J.R., Vandermeulen, R.A., Montavon, G., Samek, W., Kloft, M., Dietterich, T.G., Müller, K.-R.: A unifying review of deep and shallow anomaly detection **109**(5), 756–795 (2021) <https://doi.org/10.1109/JPROC.2021.3052449>
- Reed, I.S., Yu, X.: Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution **26**(4), 653–668 (1990) <https://doi.org/10.1109/7.59861>
- Sørensen, M., Bro, R.: Automated multi-way analysis for calibration and decomposition **23**(12), 600–615 (2009) <https://doi.org/10.1002/cem.1269>
- Schölkopf, B., Platt, J.C., Shawe-Taylor, J., Smola, A.J., Williamson, R.C.: Estimating the support of a high-dimensional distribution **13**(7), 1443–1471 (2001) <https://doi.org/10.1162/089976601750264965>

- Sakurada, M., Yairi, T.: Anomaly detection using autoencoders with nonlinear dimensionality reduction. In: Proceedings of MLSDA 2014: 2nd Workshop on Machine Learning for Sensory Data Analysis, pp. 4–148 (2014). <https://doi.org/10.1145/2689746.2689747>
- Tax, D.M.J., Duin, R.P.W.: Support vector data description **54**(1), 45–66 (2004) <https://doi.org/10.1023/B:MACH.0000008084.60811.49>
- Tabernik, D., Šela, S., Skvarč, J., Skočaj, D.: Segmentation-based deep-learning approach for surface-defect detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), pp. 1187–1196 (2019). <https://doi.org/10.1109/WACV.2019.00131>
- Tucker, L.R.: Some mathematical notes on three-mode factor analysis **31**(3), 279–311 (1966) <https://doi.org/10.1007/BF02289464>
- Zhou, C., Paffenroth, R.C.: Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 665–674 (2017). <https://doi.org/10.1145/3097983.3098052>