

# Multispectral Anomaly Detection: Comparison of sensor bands in conventional and machine learning approaches

Eva Strauss<sup>1</sup>, Dimitri Bulatov<sup>1</sup>, Edwin Deisling<sup>1</sup>, Arno Schmidt<sup>2</sup>, Peter Solbrig<sup>2</sup>, Daniel Wagner<sup>2</sup>, Alexander Gatter<sup>2</sup>, Martin Kolloch<sup>2</sup>

<sup>1</sup> Fraunhofer IOSB, 76725 Ettlingen, Germany - (name.surname)@iosb.fraunhofer.de

<sup>2</sup> Rheinmetall Electronics GmbH, 28309 Bremen, Germany - (name.surname)@rheinmetall.com

**Keywords:** Anomaly Detection, Camouflage Detection, Synthetic Imagery, Thermal Simulation, Band Selection.

## Abstract

Detecting camouflaged objects in UAV-based reconnaissance imagery is challenging, particularly beyond the visible spectrum where labeled data, required for training advanced methods, is limited. Multi-spectral sensing and synthetic thermal imagery promise improved anomaly-based camouflage detection. This work systematically compares sensor-band configurations and anomaly detection paradigms, including simulated long-wave infrared (LWIR) references, for multi-spectral camouflage detection. We construct 15 datasets of multi-spectral image stacks by not only combining real imagery, but also real and synthetic LWIR, generated via 3D scene reconstruction and heat-transfer simulation. Four detectors, ranging from statistical-based, conventional machine-learning, and deep learning, are evaluated against the image band combinations, focusing on methods suitable for near real-time operation. Results show that complementary RGB, near-infrared, and LWIR information, exploited by lightweight deep learning detectors, enables accurate multi-spectral camouflage detection, and the applied thermal simulations for synthesizing LWIR are not yet accurate or well-integrated enough to serve as beneficial reference channels, suggesting that the current network and training setup cannot yet exploit the simulated channel. This motivates for future improvement of thermal models, reduction of reality-to-simulation gaps, and tailoring deep learning architectures or training strategies to harness synthetic data.

## 1. Introduction

Many real-world tasks such as infrastructure or environmental monitoring benefit from the vast developments in machine learning (ML). Although modern ML networks achieve state-of-the-art performance in object detection and image segmentation, they often rely on large volumes of data. Considering non-visible spectra such as near- to longwave-infrared, adequate data is often difficult to obtain, and tailored object detectors are prone to fail. However, these spectra are particularly relevant in safety-related tasks, e.g. wildfire detection, urban heat studies, or military reconnaissance. The latter additionally suffers from camouflage as a countermeasure, mostly in the visible spectrum, yet research and appropriate datasets are scarce (Liu and Di, 2023). To tackle this challenge, anomaly detection may provide an alternative: rather than searching for known object classes, unexpected or unusual changes in the environment are identified. Existing approaches to anomaly detection in imaging either rely on statistical deviations, saliency assumptions, or learning some hidden (deep) features from the data. When the expected thermal state of a scene is known, anomaly detection can be naturally formulated as a change detection problem. This perspective not only enables physically interpretable and robust detection of deviations but may also provide a promising direction for integrating model-based and learning-based approaches.

In this study, the objective is to analyze best image band combinations towards the detection of camouflaged objects. Our core contribution is to extend this variety of spectra with simulated longwave images, instantiated from the enhanced 3D model of the underlying scene. The latter represents the expected state without anomalies, and thus shifts the anomaly definition from the input distribution to deviations between the reality and simulation (Burkard et al., 2020). Guided by real-life practicability in terms of sensors, platforms, and recon-

naissance operations, considering multi-spectral imagery with typically RGB channels (both full and separate RGB bands) and infrared channels (near-infrared, red-edge and longwave infrared) is reasonable. Furthermore, detector methodologies should be potentially applicable in near real-time with low processing costs. To obtain a comprehensive overview, we select one detector from each of the categories: statistics-based, conventional machine-learning, and modern deep learning.

## 2. Related Work

Early approaches to anomaly detection in images rely on statistical modeling and thresholding techniques. Assumptions are exploited that normal pixel distributions can be described using parametric models, such as Gaussian distributions or mixture models, and that anomalies correspond to outliers. In saliency-based approaches, anomalies are identified as regions that stand out from their surroundings. Detecting thermal anomalies means identifying regions that are salient in infrared images but not in corresponding visible-spectrum images (Sledz and Heipke, 2021). Such approaches implicitly assume that anomalies exhibit strong local contrast, which may not always hold in thermally homogeneous environments. Deep learning approaches significantly improved anomaly detection performance by generalizing the conventional approaches in the direction of latent features. In reconstruction-based anomaly detection, deep learning networks such as autoencoders or variational autoencoders are trained exclusively on normal data, and at inference time, anomalies are identified via high reconstruction errors, under the assumption that abnormal patterns cannot be faithfully reconstructed (Bergmann et al., 2019; Behrouzi et al., 2023). However, for deep-learning-based approaches, thermal images pose challenges not only due to their low texture and smooth gradients, but also because of non-seldom lack of adequate training data, which holds likewise for camouflage datasets. In Hwang

and Ma (2026), a military camouflage dataset of RGB imagery was generated and evaluated in its applicability by the deep learning network SINet-V2 (Fan et al., 2021). Civil camouflage datasets consisting of RGB imagery of mostly camouflaged animals were brought into the training process as well to investigate detection performance based on this additional training data, and an increase in performance was indeed observed. Detection performance of camouflaged objects as a function of input image modality (RGB, LWIR, or NIR) was studied in Laidouni et al. (2025), using the deep learning object detector YOLOv5m (Zhou et al., 2023). Results showed best performances for RGB imagery, followed by NIR images, and lowest performance on LWIR imagery and the need for investigating fused image spectra was stated. In Hupel and Stütz (2022), a new dataset for multi-spectral, UAV-based camouflage detection was presented and four anomaly detectors had been evaluated on said dataset, revealing the applicability of conventional, non-ML based methods. The authors furthermore demonstrated in Hupel and Stütz (2024) that a managed selection of three sensor bands can achieve comparable performance.

### 3. Methodology

#### 3.1 Experimental Setup

The experimental setup consists of 15 image subsets and four anomaly detection methods to be tested. The image subsets are generated from real multi-spectral imagery and synthetically generated LWIR images which are co-registered with the corresponding real LWIR data. The selected anomaly detectors embrace statistical, conventional machine-learning and modern deep learning methods. The performance in each experiment is evaluated based on the common metrics recall, precision, and F1 score. In the case of the deep-learning-based network, we additionally consider mean average precision at an intersection over union of 0.5. This is done because it returns bounding boxes, unlike other detectors with pixel-wise annotations, as output. Since object-based detection is what matters for the user, assessment of a bounding box using a meaningful, fixed threshold provides a better insight into the actual network performance than final recall and precision metrics.

In the following, the preparation of the dataset is outlined first, followed by a summary of the synthetic image generation. Eventually, an overview of the four selected detectors is given.

#### 3.2 Dataset Preparation

The dataset MUDCAD-X (Hupel and Stütz, 2023) provides nadir-view multi-spectral imagery including RGB, separated red (R), green (G), and blue (B) channels, near-infrared (NIR), red-edge (RE) and longwave infrared (LWIR) images, acquired by a multi-spectral sensor system mounted on an unmanned aerial vehicle (UAV). The images of each channel are aligned and of same size and resolution with a ground sampling distance of 0.1 m. The two recorded test areas are located on the campus of the University of the Bundeswehr Munich and show grassland, streets, and forests. Anomalies include tarps, camouflage nets, and people in both green and yellow military uniforms. Before each flight, these anomalies are placed within the scene as static objects, i.e. no movement during the flight, and their positions are documented, yielding corresponding label maps in the dataset. Measurements were executed in spring, summer, and autumn, with several flights per day yielding 853 images in total.

From the dataset, we derive new subsets of image arrangements for our given objective. First, images are stacked into three-channel imagery, as illustrated in Figure 1, to provide the following combinations: RGB, HSV, NIR–G–B, RE–G–B, NIR, NIR–RE–LWIR, NIR–GS–LWIR, NIR–LWIR, LWIR, S–LWIR, S–LWIR–NIR, S–LWIR–RE, S–LWIR–H, S–LWIR–GS, S–NIR–RE, where GS is the RGB image converted to grayscale, H refers to the Hue channel of the HSV-converted RGB image, and S states the synthetic reference image, generated as outlined in the subsequent Sec. 3.3. The named combinations are selected based on their potential availability in real life scenarios and their varying characteristics. In detail, RGB and HSV highlight the visible-spectrum texture and color contrast for general scene interpretation and man-made structure delineation. NIR and RE are most feasible to capture vegetation vigor due to the reflection properties in these wavelengths. Finally LWIR introduces the thermal spectrum whose characteristics differ entirely from the visible to NIR spectrum as LWIR signatures are defined predominantly by thermal emission. In total, each subset exhibits eventually 272 image stacks. To promote generalization, the given labels were summarized to one label, namely the anomaly class.

#### 3.3 Synthetic Reference Image Generation

The simulated images are generated following Bulatov et al. (2020) and Strauß and Bulatov (2024) with slight simplifications within the thermal model. Using an open-source digital surface model (DSM), digital terrain model (DTM), and orthophoto as starting point the area of the test site is first reconstructed as 3D triangular mesh, see Figure 2. Within the process, land cover classification is carried out following Qiu et al. (2022), and land cover classes are matched to the thermal simulation setup, i.e. the following classes are considered: asphalt, grassland, trees, bushes, and forest boxes. The latter result from single tree detection and fusion towards a box-based forest model, simplifying the subsequent heat simulation. Within the thermal simulation setup, material characteristics are provided according to the given class label and include the materials density, heat capacity, heat conductivity, thermal emissivity, and solar albedo. Particularly, each class label offers not only the surface materials characteristics, but also sub-surface approximations such as soil below grass or asphalt starting at a determined depth. The named material parameters are fetched from a material database within the simulation framework, whereas the data therein is based on an extensive literature review (Bulatov et al., 2020).

Given the reconstructed 3D triangular mesh including material properties, the thermal simulation is carried out based on the general heat transfer equation. Each surface triangle is virtually expanded along its normal direction to form a prism, being discretized for numeric simulation into layers of depth  $d$ , in order to model conductive heat dissipation into the ground or interior. At the surface, convective and radiative heat exchange occur. With the subscript  $s$  referring to the surface material properties and  $i$  referring to the respective material of the  $i$ -th sub-surface layer in the prism, the heat transfer equations can be written as (Strauß and Bulatov, 2024)

$$\begin{aligned} \rho_s c_{v,s} d \frac{\partial T_s}{\partial t} = & \frac{k_s}{d} (T_1 - T_s) \\ & + h (T_{\text{air}} - T_s) \\ & + \gamma_{\text{sky}} \sigma_B \varepsilon_s (T_{\text{sky}}^4 - T_s^4) \\ & + \gamma_{\text{sun}} (1 - a) E_s \cos(\phi) \end{aligned} \quad (1)$$

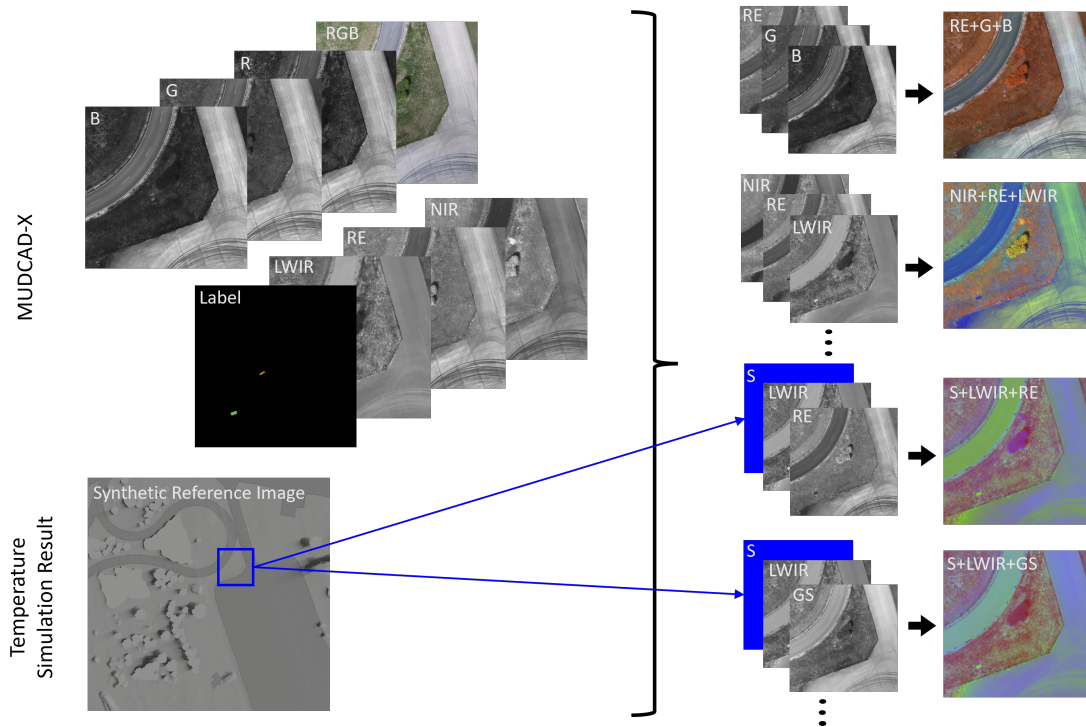


Figure 1. Preparation of the dataset. Top left: examples of the MUDCAD-X dataset. Bottom left: result of the simulation result to create synthetic reference images. Right: Examples of the stacked one-channel images.

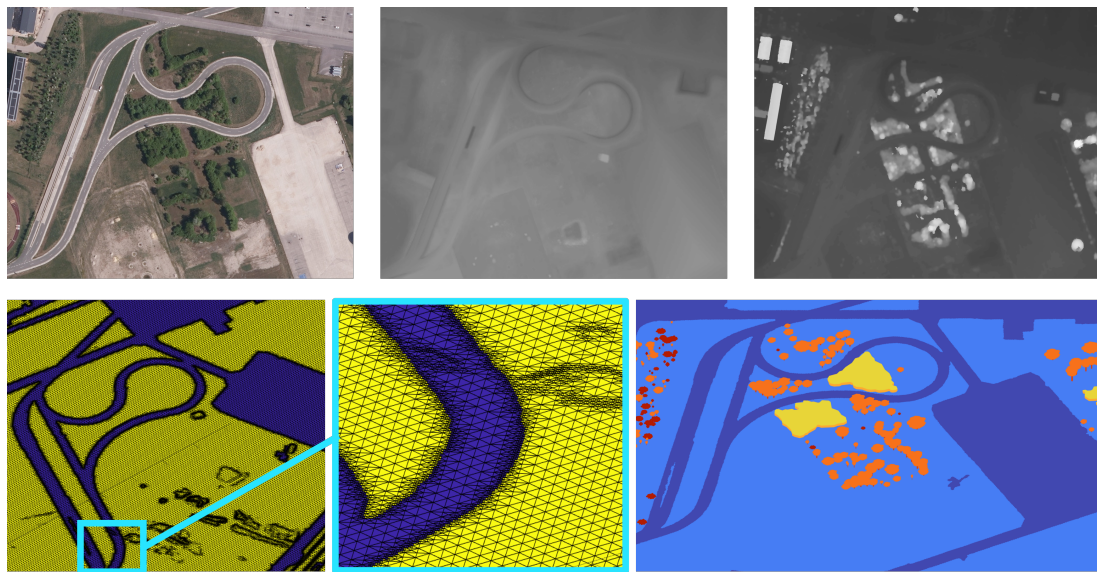


Figure 2. Input and output of the 3D reconstruction for thermal simulation. Top: Orthophoto (left), DTM (middle), DSM (right). Bottom: Ground mesh with detail (left and middle), 3D mesh with color-coded material classification (right) including asphalt street (dark blue), grassland (light blue), forest boxes (yellow), trees (orange) and bushes (red).

and

$$\rho_i c_{v,i} d \frac{\partial T_i}{\partial t} = \frac{k_{i-}}{d} (T_{i-1} - T_i) + \frac{k_{i+}}{d} (T_{i+1} - T_i), \quad (2)$$

with the temperature of the surface, first or  $i$ -th layer  $T_{\{s,i,1\}}$  respectively, the material density  $\rho_{\{s,i\}}$ , specific heat capacity  $c_{v,\{s,i\}}$ , and heat conductivity  $k_{\{s,i\}}$  each of surface or  $i$ -th layer respectively, the uniform prism layers depth  $d$ , convective heat transfer coefficient  $h$ , air temperature  $T_{\text{air}}$  and sky tem-

perature  $T_{\text{sky}}$ , solar occlusion and sky visibility factors  $\gamma_{\text{sun}}$  and  $\gamma_{\text{sky}}$ , the Boltzmann constant  $\sigma_B$ , thermal emissivity  $\varepsilon_s$ , solar albedo  $a$ , solar irradiation at the earth's surface  $E_s$ , and the relative surface orientation angle  $\phi$ . The heat conductivity parameters  $k_{i-}$  and  $k_{i+}$  refer to the corresponding boundary conductivities of the  $i$ -th layer. The convective heat transfer coefficient is modeled empirically based solely on the predominant wind speed. The sky temperature is modeled based on air temperature, dew-point and hour of the day, cp. Lienhard IV and Lienhard V (2019). Along with air temperature, the parameters

wind speed, humidity, and cloud cover, each needed within  $E_s$  (Duffie and Beckman, 1991),  $T_{sky}$ , or  $h$ , are drawn from open-source weather server data utilizing the *meteostat* python API (Met, 2026).

The thermal simulation results are given as temperature values per triangle in the 3D mesh. To achieve simulated 2D imagery aligned with the real MUDCAD-X images, the full area is first rendered from a nadir view (orthogonal projection) yielding a synthetic thermal orthophoto. Second, the orthophoto is split into the according image frames to match the real data. Due to the relatively small size of the test area, a sufficient manual approach was chosen.

### 3.4 Applied Detectors

To evaluate anomaly detection based on band-selection and simulated data, four detectors are employed which vary in their fundamental design.

First, the statistics-based **Reed-Xiaoli Detector (RXD)** by Reed and Yu (1990) is applied. The RXD defines anomalies by thresholding the squared Mahalanobis Distance of each pixel in an image  $I$ ,

$$\delta_{RXD}(p_I) = \left[ (p_I - \mu_I)^T \Sigma_I^{-1} (p_I - \mu_I) \right] > 3\sigma_{RXD}, \quad (3)$$

with the pixel under test  $p_I$  of the image  $I$ , the image mean  $\mu_I$ , the covariance matrix  $\Sigma_I$  of image  $I$ , and the standard deviation  $\sigma_{RXD}$  of  $\delta_{RXD}(I)$ . We set the threshold to differ between normality and anomaly to three standard deviations of the RXD score. The RXD is known to be robust to noise and sensitive to subtle anomalies in hyperspectral imagery. Since a Gaussian background is assumed within the RXD, however, it can yield to a high false alarm rate in highly variable scenes.

As second and third method, we investigate the conventional ML-based **Isolation Forest (IF)** by Al Farizi et al. (2021), in both pixel- and superpixel-wise manner. The IF handles high-dimensional data well but can struggle with conditional anomalies due to random feature selection and splits, as anomalies are defined by dynamic forest construction and thresholding the split count. In detail, a forest of isolation trees is build upon the given input pixels of an image, where each tree is basically a series of random binary splits until instances are isolated, i.e. leafs with a single instance. The anomalous pixel score is then determined based on split count and path length within the isolation forest,

$$\delta_{IF}(p_I) = 2^{\frac{E(h(p_I))}{c(n)}}, \quad (4)$$

with the isolation depth  $h(p_I)$  of the pixel  $p_I$ , the average isolation depth across all isolation trees  $E(h(p_I))$ , the expected depth of a randomly chosen isolation tree  $c(n)$ , and the number  $n$  of pixels in  $I$ . The threshold is set to the 0.95 quantile of the IF scores. The IF can be applied in terms of novelty detection where preceding training based on non-anomalous data is required, or in terms of outlier detection with no training needed. In this study, we apply the latter to avoid further data processing and maintain comparability to the RXD. Generally, the generation of the isolation forest depends on the given input features, thus this methodology relies on feature engineering. Therefore, preceding to the IF application, we analyze the pixel- and superpixel-wise features of the MUDCAD-X dataset beforehand.

Finally, the deep learning network **You-only-look-once, version 10 nano, (YOLOv10n)**, see Wang et al. (2024), is applied as fourth detector. Even though being an object detector and not an anomaly detector, YOLOv10n was chosen for its easy adaptability, fine-tuning capabilities and real-time usability as being a lightweight version of the well-known YOLO network. Also, YOLO had already demonstrated its feasibility in several related tasks (Laidouni et al., 2025; Pichler and Hueber, 2024; Klein et al., 2024). Furthermore, deep learning approaches hold the greatest promise because they can automatically learn complex feature representations and potentially discern subtle patterns that distinguish anomalies from the background. In particular, they benefit from hierarchical feature extraction, enabling them to uncover intricate relationships within the image data that traditional methods might miss. Given their capacity to learn from large datasets and improve with training, deep learning is considered a promising path to enhance anomaly detection performance on the MUDCAD-X dataset. To tailor the network to the given task, the pretrained version of YOLOv10n is configured to detect one class only, namely the *anomaly* class, and subsequently fine-tuned using the prepared image stack datasets. Whilst the pretraining had been based on the COCO dataset (Lin et al., 2014) with hundreds of thousands images including 80 classes and solely RGB images, our dataset consists of 272 images and is divided into training, validation, and test sets, with 90% allocated to training and 10% each to validation and testing.

## 4. Results

### 4.1 Feature Space Analysis

Before application of the four anomaly detectors, a preliminary feature analysis of the MUDCAD-X dataset is conducted. Figure 3 displays the scatterplot of 12 image features: each separate RGB channel, RE, NIR and LWIR channels, the full RGB orthophoto, each HSV channel (hue, saturation and value), and the normalized digital vegetation index (NDVI) as well as entropy. The NDVI, calculated from NIR and red bands to quantify vegetation health and density, is chosen as it reveals vegetation vigor and distribution in an image, potentially allowing to separate natural background from the man-made anomalies in the MUDCAD-X dataset. Image entropy measures the randomness or variability of pixel values, revealing texture complexity and information content in an image.

The observed feature subspaces reveal a challenge in separating background from anomalies, as anomalous data points are entirely enveloped by background points. This observation suggests that image-based anomaly detection using conventional methods is likely to struggle. Consequently, some feature engineering is necessary to adapt the Isolation Forest approach to this data. This is carried out by simple linear iterative clustering (SLIC, (Achanta et al., 2010)) of the images in order to partition the input image stack into homogeneous regions based on pixel similarity. For feature engineering, each superpixel is assigned with statistics such as mean value, variance, and most frequent value. Analyzing the resulting superpixel-based feature subspaces across image stacks now shows better separation between anomalies and background, see the respective feature subspaces of the LWIR and NIR images in Figure 4 and Figure 5. Although anomalous pixels can reside within background regions in the subspace representation, the overall distributions for the two classes differ, for example in skewness or kurtosis.



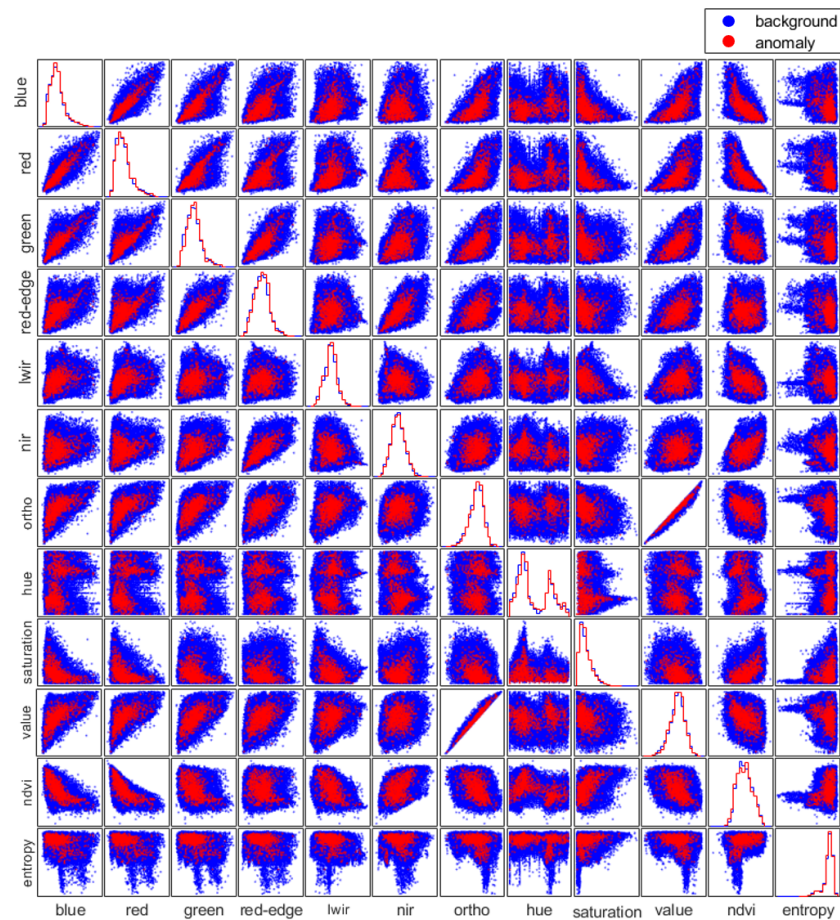


Figure 3. Pair-wise scatterplot of the feature-space of the MUDCAD-x dataset. The plots along the diagonal show the aggregated histograms for each feature. Each point within the pair-wise plots corresponds to a single pixel from a MUDCAD-X image, labeled as background or anomaly according to pixel-wise annotations, with all dataset images included.

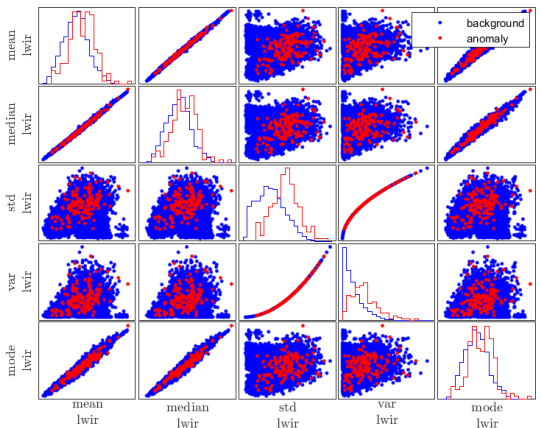


Figure 4. Scatterplot of the superpixel-wise feature subspace of the longwave infrared imagery in the MUDCAD-X dataset.

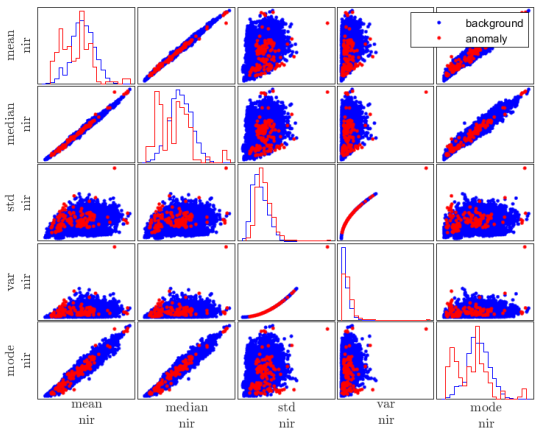


Figure 5. Scatterplot of the superpixel-wise feature subspace of the near infrared imagery in the MUDCAD-X dataset.

## 4.2 Anomaly Detection

**Reed-Xiaoli Detector** The RXD is run separately on each image stack, with outputs illustrated in Table 2 and full findings summarized in Table 1. In summary, the RXD emerged to be unsuitable for the given task yielding overall low performance

metrics, even though the RXD had already proven to be of acceptable performance in Hupel and Stütz (2024). Therein, band selection was achieved by a managed approach depending on environmental context, yet final band combinations were not reported. This underscores the need for both further investigation of wavelength-band interplay and potentially deeper fea-

Table 1. Evaluation metrics per band-set, with best results per detector marked in bold.

| Image Stack     | RXD          |              |              | IF pixel-wise |              |              | IF superpixel-wise |              |              | YOLOv10n     |              |              |              |
|-----------------|--------------|--------------|--------------|---------------|--------------|--------------|--------------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                 | F1           | Precision    | Recall       | F1            | Precision    | Recall       | F1                 | Precision    | Recall       | F1           | MaP          | Precision    | Recall       |
| RGB             | 0.113        | 0.074        | <b>0.243</b> | <b>0.107</b>  | <b>0.059</b> | <b>0.552</b> | 0.053              | 0.030        | 0.269        | 0.964        | 0.977        | 0.983        | 0.945        |
| HSV             | 0.055        | 0.046        | 0.068        | 0.068         | 0.038        | 0.354        | 0.076              | 0.042        | 0.374        | 0.956        | 0.970        | 0.980        | 0.933        |
| NIR – G – B     | 0.064        | 0.043        | 0.133        | 0.086         | 0.047        | 0.447        | 0.065              | 0.036        | 0.304        | 0.966        | 0.979        | 0.991        | 0.942        |
| RE – G – B      | 0.101        | 0.067        | 0.207        | 0.084         | 0.047        | 0.435        | 0.063              | 0.035        | 0.286        | 0.965        | 0.971        | <b>0.999</b> | 0.933        |
| NIR             | 0.065        | <b>0.160</b> | 0.040        | 0.010         | 0.005        | 0.247        | 0.089              | 0.050        | 0.455        | 0.962        | <b>0.984</b> | 0.991        | 0.933        |
| NIR – RE – LWIR | 0.075        | 0.049        | 0.163        | 0.049         | 0.027        | 0.255        | 0.100              | 0.055        | 0.517        | 0.955        | 0.967        | 0.974        | 0.937        |
| NIR – GS – LWIR | 0.049        | 0.034        | 0.089        | 0.043         | 0.024        | 0.224        | <b>0.082</b>       | 0.045        | 0.450        | <b>0.982</b> | 0.983        | 0.998        | <b>0.967</b> |
| NIR – LWIR      | <b>0.128</b> | 0.135        | 0.121        | 0.034         | 0.019        | 0.179        | <b>0.106</b>       | <b>0.059</b> | <b>0.566</b> | 0.974        | 0.979        | 0.998        | 0.950        |
| LWIR            | 0.038        | 0.024        | 0.085        | 0.011         | 0.006        | 0.278        | 0.036              | 0.020        | 0.190        | 0.972        | 0.970        | 0.991        | 0.953        |
| S – LWIR        | 0.037        | 0.025        | 0.072        | 0.014         | 0.008        | 0.072        | 0.032              | 0.018        | 0.175        | 0.882        | 0.891        | 0.943        | 0.829        |
| S – LWIR – NIR  | 0.048        | 0.032        | 0.097        | 0.068         | 0.038        | 0.353        | 0.079              | 0.044        | 0.409        | 0.933        | 0.957        | 0.964        | 0.903        |
| S – LWIR – RE   | 0.053        | 0.036        | 0.098        | 0.075         | 0.042        | 0.388        | 0.081              | 0.045        | 0.410        | 0.950        | 0.954        | 0.982        | 0.920        |
| S – LWIR – H    | 0.074        | 0.056        | 0.112        | 0.053         | 0.029        | 0.275        | 0.075              | 0.041        | 0.373        | 0.857        | 0.874        | 0.940        | 0.787        |
| S – LWIR – GS   | 0.019        | 0.013        | 0.038        | 0.033         | 0.018        | 0.172        | 0.051              | 0.028        | 0.250        | 0.932        | 0.954        | 0.973        | 0.894        |
| S – NIR – RE    | 0.066        | 0.044        | 0.135        | 0.079         | 0.044        | 0.407        | 0.079              | 0.044        | 0.396        | 0.933        | 0.939        | 0.969        | 0.900        |

ture engineering. Highest F1-Score (12.8%) is achieved by combining NIR and LWIR bands, i.e. combining vegetative and temperature characteristics and thus those features in which the anomalies (man-made camouflage) potentially differ most from their natural surroundings. Though NIR as standalone input yields still higher precision than other bands (16.6%), the RXD fails tremendously at the LWIR band, both as standalone and when combined with the synthetic reference image. Qualitatively, the RXD results tend to spot extreme-value regions (hotspots), which does not align with usual camouflage characteristics, therefore yielding low performance metrics. When the simulated reference is used within the image stacks, some false positives arise from simulation issues like misalignment and unrepresented ground variation, however, they overshoot the expected false positive rate, revealing RXD's difficulty to deal with simulated data.

**Isolation Forest** The IF is first applied in a pixel-wise manner, i.e. each channel of the image stacks can be interpreted as an individual image feature. The severely low values per each evaluation metric, cp. Table 1, third column, show that the pixel-level application suffers from overlap in feature spaces and performs poorly on the MUDCAD-X data. Best performance is achieved on RGB imagery only with a clear distance to other image stacks, yet only about half the anomalies are detected, along with false positives. Once IF is applied to superpixels, the number of false positives drops and precision improves markedly compared with the pixel-wise approach for most band combinations, though recall only rises for stacks without RGB channels. The strongest performance is achieved with combined NIR and LWIR bands, as did the RXD in terms of F1-Score. Yet, evaluation metrics on NIR-LWIR image stacks with the superpixel approach perform similarly to pixel-wise IF on RGB imagery, indicating the necessity of band-tailored anomaly detectors.

**YOLOv10n** Adapting and fine-tuning the pretrained deep-learning-based detector YOLOv10n clearly outperformed RXD and IF across all metrics, cp. Table 1. Recall is maximized when using the multi-channel input composed of NIR, grayscale RGB, and LWIR data, and this combination also achieves strong F1-Score and mAP. This setup combines diverse lighting conditions (sunlight, residual light from dawn, dusk, or moonlight) and various object characteristics, such as natural vegetation versus manmade materials, color, shape, and temperature. Even without RGB information, combining NIR and LWIR still yields high performance. LWIR alone also performs very strongly, and using RGB or NIR alone provides comparable results to the three-channel configurations. This is noteworthy for two reasons: first, LWIR provides the feasibility for camouflage detec-

tion during night, independently of external lighting. Second, these findings enable flexibility during operation, as reliance on specific bands and corresponding operational sensors is reduced. Introducing the synthetic reference image generally reduces performance, with the most noticeable decline occurring for the mixed simulated and real LWIR data. This indicates that, within the current architecture and training configuration, the model is widely unable to exploit this channel in a beneficial manner. Possible reasons include insufficient simulation accuracy as well as the potential necessity to adjust the model towards synthetic data, either by architectural changes or hyperparameter optimization.

## 5. Conclusions

Across methods and modality combinations, the NIR, LWIR, and RGB-derived information consistently yielded the strongest results, aligning with their complementary physics (reflectance vs. emission) and discriminative power for natural vs. man-made materials. Incorporating simulated references did not improve detection, potentially because the anomaly detectors had not been tuned towards synthetic data, and because the overall reality-to-simulation gap remains too large. Future work should expand across several lines of research: first, narrowing the reality gap via enhanced thermal modeling, richer scene reconstruction, and refined sensor simulation should be tackled. Second, RXD's sensitivity to simulation errors suggests a possible role as a tool for diagnosing and improving simulation components or informing downstream detectors, similar to Burkard et al. (2020) and Strauß and Bulatov (2022). Third, the extraordinary performance of the YOLO network should be compared against tailored DL anomaly detection networks or DL camouflage detection networks, particularly focusing on approaches that do not rely on training data of the defined anomaly or camouflaged object.

## References

- Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., Süsstrunk, S., 2010. SLIC superpixels. Technical report, Technical Report EPFL.
- Al Farizi, W. S., Hidayah, I., Rizal, M. N., 2021. Isolation forest based anomaly detection: A systematic literature review. *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*, IEEE, 118–122.
- Behrouzi, S., Dix, M., Karampanah, F., Ates, O., Sasidharan, N., Chandna, S., Vu, B., 2023. Improving visual defect detec-

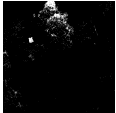
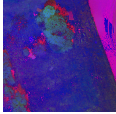
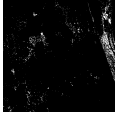
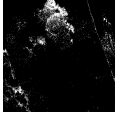
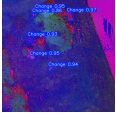
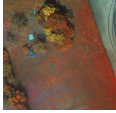
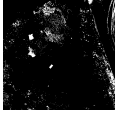
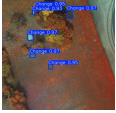
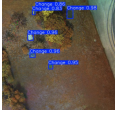
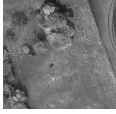
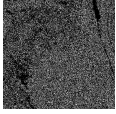
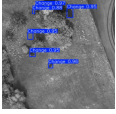
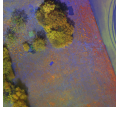
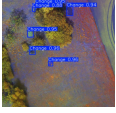
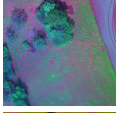
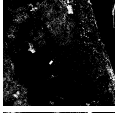
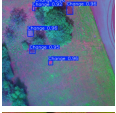
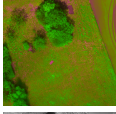
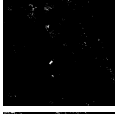
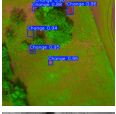
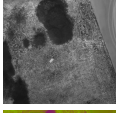
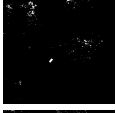
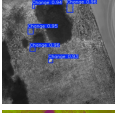
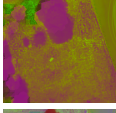
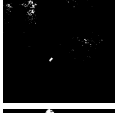
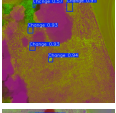
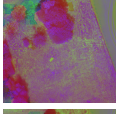
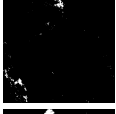
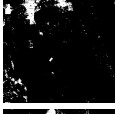
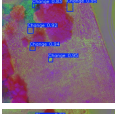
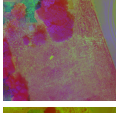
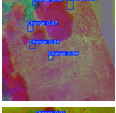
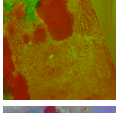
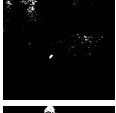
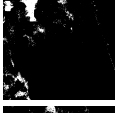
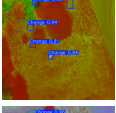
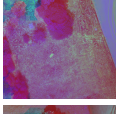
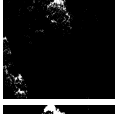
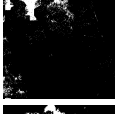
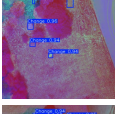
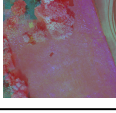
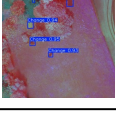
| Image Stack     | Sample                                                                              | RXD                                                                                 | IF pixel-wise                                                                       | IF superpixel-wise                                                                    | YOLOv10n                                                                              |
|-----------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
| RGB             |    |    |    |    |    |
| HSV             |    |    |    |    |    |
| NIR – G – B     |    |    |    |    |    |
| RE – G – B      |    |    |    |    |    |
| NIR             |    |    |    |    |    |
| NIR – RE – LWIR |    |    |    |    |    |
| NIR – GS – LWIR |   |   |   |   |   |
| NIR – LWIR      |  |  |  |  |  |
| LWIR            |  |  |  |  |  |
| S – LWIR        |  |  |  |  |  |
| S – LWIR – NIR  |  |  |  |  |  |
| S – LWIR – RE   |  |  |  |  |  |
| S – LWIR – H    |  |  |  |  |  |
| S – LWIR – GS   |  |  |  |  |  |
| S – NIR – RE    |  |  |  |  |  |

Table 2. Examples of the anomaly detection results on each generated image stack.

- tion and localization in industrial thermal images using autoencoders. *J. Imaging*, 9(7/137).
- Bergmann, P., Fauser, M., Sattlegger, D., Steger, C., 2019. Mvtec ad – a comprehensive real-world dataset for unsupervised anomaly detection. *IEEE Conference on Computer Vision and Pattern Recognition*, 9592–9600.
- Bulatov, D., Burkard, E., Ilehag, R., Kottler, B., Helmholz, P., 2020. From multi-sensor aerial data to thermal and infrared simulation of semantic 3D models: Towards identification of urban heat islands. *Infrared Physics & Technology*, 105, 103233.
- Burkard, E., Bulatov, D., Kottler, B., 2020. Towards detection of thermal anomalies in large urban areas using simulation. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43, 1195–1202.
- Duffie, J. A., Beckman, W. A., 1991. *Solar engineering of thermal processes*. 2 edn, John Wiley.
- Fan, D.-P., Ji, G.-P., Cheng, M.-M., Shao, L., 2021. Concealed object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10), 6024–6042.
- Hupel, T., Stütz, P., 2022. Adopting hyperspectral anomaly detection for near real-time camouflage detection in multispectral imagery. *Remote Sensing*, 14(15), 3755.
- Hupel, T., Stütz, P., 2023. Measuring and predicting sensor performance for camouflage detection in multispectral imagery. *Sensors*, 23(19), 8025.
- Hupel, T., Stütz, P., 2024. Sensor-managed anomaly detection for camouflage detection in airborne multispectral imagery. *IEEE Aerospace Conference*, IEEE, 1–11.
- Hwang, K.-S., Ma, J., 2026. Military camouflaged object detection with deep learning using dataset development and combination. *The Journal of Defense Modeling and Simulation: Applications, Methodology, Technology*, 23(1), 67–78.
- Klein, M. D., Edel, Z. J., Packard, C. D., Hendrickson, J. N., Levanen, A. C., Rynes, P. L., 2024. Automatic detection and recognition of humans with yolo-based deep learning using muses-generated eo/ir synthetic imagery. *Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques, and Applications II*, SPIE, 12.
- Laidouni, M. Z., Bondžulić, B., Pavlović, B., 2025. Military camouflaged target detection through multimodal images: A comparative study. *International Scientific Conference on Military Sciences - VojNa 2025*.
- Lienhard IV, J., Lienhard V, J., 2019. *A Heat Transfer Textbook*. 4 edn, Dover Publications, Inc.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C. L., 2014. Microsoft coco: Common objects in context. *European conference on computer vision*, Springer, 740–755.
- Liu, M., Di, X., 2023. Extraordinary MHNet: Military high-level camouflage object detection network and dataset. *Neurocomputing*, 549, 126466.
- Met, 2026. Meteostat developers. <https://dev.meteostat.net/>. Accessed: 2026-03-17.
- Pichler, A., Hueber, N., 2024. Method for training deep neural networks in vehicle detection using drone-captured data and background synthesis. *Synthetic Data for Artificial Intelligence and Machine Learning: Tools, Techniques, and Applications II*, 13035, SPIE, 175–187.
- Qiu, K., Budde, L. E., Bulatov, D., Iwaszczuk, D., 2022. Exploring fusion techniques in U-Net and DeepLab V3 architectures for multi-modal land cover classification. *Earth Resources and Environmental Remote Sensing/GIS Applications XIII*, 12268, SPIE, 190–200.
- Reed, I., Yu, X., 1990. Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 38(10), 1760–1770.
- Sledz, A., Heipke, C., 2021. Thermal anomaly detection based on saliency analysis from multimodal imaging sources. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 1, 55–64.
- Strauß, E., Bulatov, D., 2022. A region-based machine learning approach for self-diagnosis of a 4D digital thermal twin. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 10, 265–272.
- Strauß, E., Bulatov, D., 2024. Between gaming and microclimate simulations: Temperature estimation of an urban area. *International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications 2024*, 70–80.
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J. et al., 2024. YOLOv10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems*, 37, 107984–108011.
- Zhou, F., Deng, H., Xu, Q., Lan, X., 2023. CNTR-YOLO: improved YOLOv5 based on ConvNext and transformer for aircraft detection in remote sensing images. *Electronics*, 12(12), 2671.