

Detection of Unknown Substances in Operation Environments Using Multispectral Imagery and Autoencoders

Peer Schütt^{a,*}, Jonas Grzesiak^b, Christoph Geiß^b and Tobias Hecking^a

^aInstitute of Software Technology, German Aerospace Center (DLR)

^bInstitute of Technical Physics, German Aerospace Center (DLR)

ORCID (Peer Schütt): <https://orcid.org/0000-0002-6513-5235>, ORCID (Jonas Grzesiak): <https://orcid.org/0000-0001-9690-0780>, ORCID (Christoph Geiß): <https://orcid.org/0000-0001-6518-0012>, ORCID (Tobias Hecking): <https://orcid.org/0000-0003-0833-7989>

Abstract. Autonomous vehicles and robotic systems are increasingly used to perform operations in environments that bear potential risks to humans (e.g. areas affected by natural disasters, warfare, or planetary exploration). One source of danger is the contamination with hazardous substances. In order to improve situational awareness and planning, such substances must be detected using the sensors of the autonomous system. However, training a supervised machine learning model to detect different substances requires a labelled dataset with all potential substances to be known in advance, which is often impracticable. A possible solution for this is to pose an anomaly detection problem where an unsupervised algorithm detects suspicious substances that differ from the normal operation environment.

In this paper we propose SpectrAE, a convolutional autoencoder-based system that processes multispectral imaging data (covering visible to near-infrared ranges) to identify surface anomalies on roads. Unlike traditional detection methods such as gas chromatography and physical sampling that risk contamination and cause operational delays, or laser-based remote sensing techniques that require pre-localisation of potential hot spots, our approach offers near real-time detection capabilities without prior knowledge of specific hazardous substances. The system is trained exclusively on normal road conditions and identifies potential hazards through localised reconstruction loss patterns, generating Areas of Interest for further investigation. Our contributions include a robust end-to-end detection pipeline, comprehensive evaluation of system performance, and a roadmap for future development in this emerging intersection of autonomous systems and crisis response technologies.

1 Introduction

Operations in uncertain environments face critical challenges, for example, transporting relief supplies through hazardous and potentially contaminated areas. To this end, remotely controlled and autonomous vehicles are increasingly used to reduce the risk for humans. This comes with the need for technical solutions regarding environmental perception, hazard assessment, and route navigability among other challenges.

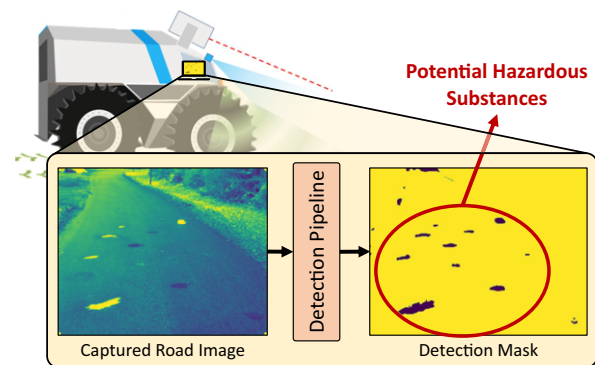


Figure 1. End-to-end illustration of our substance detection pipeline. The system processes raw multispectral input images (left), extracts relevant spectral features through an autoencoder architecture, and generates a binary detection mask (right). These masks enable subsequent analysis and characterisation of the substances within the regions.

In this work, we address a key open problem: the autonomous identification of unknown and potentially hazardous substances along roads. A primary obstacle is the lack of comprehensive, labelled datasets for supervised learning, as most hazardous substances are rare or unknown a priori. We therefore frame the task as an anomaly detection problem, flagging substances that differ from the established environmental norms as potential hazards.

Traditional methods, such as gas chromatography or swipe sampling for biological substances, involve the risk of contamination of personnel as they require physical contact with the substance. Instantaneous recognition, and thus time-critical path planning, are almost impossible with these methods. Established remote sensing methods—like laser spectroscopy—allow for safer, stand-off detection but often scan only small, targeted regions at a time. Consequently, a crucial prerequisite is the reliable identification, from a distance, of potential hot spots on road surfaces where contamination may have occurred.

To address this, we propose a new machine learning pipeline for anomaly detection on road surfaces, leveraging multispectral imaging and an unsupervised convolutional autoencoder [7] architecture, which we call SpectrAE. The system is first trained on multispec-

* Corresponding Author. Email: peer.schuett@dlr.de.

tral images representing normal road conditions, learning to accurately reconstruct these inputs. When subsequently exposed to environments containing unknown substances on roads, the model identifies these anomalies through localised reconstruction loss patterns, generating Areas of Interest (AoIs) for further analysis. This workflow is illustrated in Figure 1.

The focus of this paper is to demonstrate the performance of the SpectrAE pipeline in detecting AoIs corresponding to substances intentionally applied to road surfaces.

To the best of our knowledge, this is the first work to apply autoencoder-based anomaly detection to multispectral road imagery for autonomous hazard detection. Specifically, our contributions are:

- A novel machine learning pipeline capable of detecting unknown substances on road surfaces using multispectral imaging and unsupervised anomaly detection;
- An analysis of the detection performance of our pipeline;
- Insights and guidelines for future research, informed by the development and challenges we encountered.

Overall, we see this work as a first step towards practical remote hazard detection for autonomous crisis response vehicles. The code is available on GitHub: <https://github.com/DLR-SC/SpectrAE>.

2 Related Work

Extensive research has explored the application of autoencoder-based architectures for anomaly detection in RGB imagery [21, 2, 5, 14, 4, 28]. Typically, these approaches investigate diverse encoder designs and utilise reconstruction error as a means of identifying anomalies—a strategy also central to our proposed method.

Beyond RGB imagery, both multispectral and hyperspectral imagery have been used for anomaly detection tasks. Multispectral images contain a limited number of spectral bands, whereas hyperspectral images offer hundreds of narrow spectral bands. However, this spectral richness comes with a significant trade-off: hyperspectral information requires considerably longer acquisition times. In contrast, multispectral data can be captured rapidly.

Compared to RGB and hyperspectral imagery, anomaly detection in multispectral imagery remains relatively under-explored. Most existing multispectral anomaly detection studies rely on data gathered from satellite imagery or unmanned aerial vehicles (UAVs). For example, Lai et al. [13] employed a convolutional autoencoder framework to distinguish between coal and gangue in multispectral images in a laboratory environment. Hupel et al. [10] adapted classical anomaly detection algorithms, initially designed for hyperspectral datasets—such as the Reed-Xiaoli (RX) algorithm [19]—for camouflage detection tasks in military scenarios. Recently, Hikuwai et al. [9] applied mask region-based convolutional neural networks on multispectral satellite data to detect hazardous asbestos material in suburban residential areas. Despite these advancements, multispectral anomaly detection using ground vehicle-based platforms has notably remained absent from prior literature. Our work fills this gap, extending multispectral anomaly detection techniques into a new operational domain with significant practical implications for crisis response and autonomous navigation.

Hyperspectral imagery, on the other hand, dominates the literature on spectral anomaly detection. The appeal of hyperspectral data is partly due to its richer spectral information, as well as the widespread availability of datasets from satellite-based platforms. Classical hyperspectral anomaly detection algorithms include the RX approach [19] and statistical density estimation techniques [20, 16].

Recent developments have introduced machine learning advances, including isolation forests [24] and numerous autoencoder architectures [29, 15, 25, 6, 27, 3, 1], achieving progress towards robust anomaly detection in hyperspectral contexts.

3 Method

This section outlines our methodological approach. We begin by defining the task and describing the data acquisition process. Furthermore, we present the full machine learning pipeline, detailing each component.

3.1 Task Definition

The goal of our pipeline is to automatically detect AoIs in multispectral images that contain unknown and potentially hazardous substances on roads using vehicle-mounted sensors (Fig. 2). Multispectral cameras capture environmental data to identify potentially hazardous substances in the vehicle's path, generating AoIs as a prerequisite for targeted probing, for example, using laser-induced fluorescence (LIF) technology. The general concept is illustrated in Figure 1.

Hazardous substance detection using multispectral imaging from mobile platforms remains an under-explored research domain (see Sec. 2), necessitating our exploratory approach. Several fundamental challenges characterise this problem:

1. The concept of "hazardous substances" lacks precise definition and encompasses a highly heterogeneous class of substances.
2. No comprehensive database exists for the training of supervised detection systems.
3. Creating manually labelled datasets with sufficient scope and diversity would be very resource-intensive.

These constraints direct our investigation towards unsupervised learning methods. Our approach leverages a key insight: road surfaces exhibit high structural and compositional consistency across diverse geographic settings, providing a stable baseline for anomaly detection. By training a model to understand the characteristic spectral signatures of typical road environments, we can identify deviations that potentially indicate hazardous substance presence.

3.2 Sensor Array

Our experimental platform utilises a multi-sensor array mounted on a test vehicle (Fig. 2). While the complete array incorporates a laser-based UAV classification system (LUCS) [12], radar, an alignment laser, GPS, multispectral cameras, and brightness sensors, the presented work focuses exclusively on the multispectral imaging capabilities. Future work will integrate data from the complementary sensors to enhance detection performance. The sensor array was positioned to capture the road surface directly in the vehicle's path.

The multispectral imaging subsystem employs two specialised cameras from the SILIOS Technologies CMS series¹: the CMS-C operating in the visible spectrum (VIS, 430-700 nm) and the CMS-S operating in the near-infrared spectrum (NIR, 650-930 nm). In the following, these two sensors are referred to as VIS and NIR respectively. The cameras produce raw images at 1280 × 1024 pixel resolution. To yield multispectral images, each camera incorporates a Bayer-like mosaic filter array. Instead of a common RGB filter array,

¹ <https://www.silios.com/cms-series>

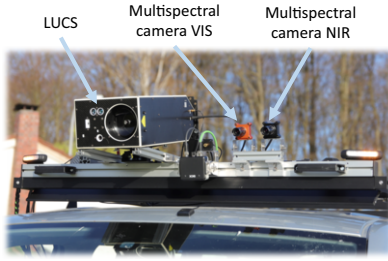


Figure 2. The current sensor array mounted on top of the test vehicle.

the array mosaic is built from 3×3 filter macro-pixels. These 3×3 filter matrices consist of 8 distinct spectral filters plus one panchromatic (greyscale) channel. The centre wavelengths of the spectral filters are approximately evenly distributed over the covered spectrum and have an average spectral resolution (full width at half-maximum) of 40 nm. The pixel values are digitised at 16-bit precision.

The cameras are aligned to have a focal point at a distance of 10 m (which is the working distance of the laser spectroscopy-based detection system). This results in a field of view of $5.7 \times 4.5 \text{ m}^2$. These raw images undergo demosaicing to separate the spectral bands, yielding 9 distinct spectral images per camera at 426×339 pixel resolution each. The cameras are run at 17 frames per second, with a customised auto-exposure algorithm. This auto-exposure ensures a maximum of 1.5% of the pixels to be saturated within a core region of the picture (687×687 pixels around the image centre). Unless otherwise specified, NIR images are used in the figures.

3.3 Data Capture

All experimental data were acquired using the vehicle-mounted sensor array described in Section 3.2 and depicted in Figure 2.

We developed a controlled experimental protocol that uses environmentally safe placeholder substances that have visually and spectroscopically similar characteristics to certain hazardous substances. These serve as proxies for hazardous substances in our anomaly detection framework. They are presented in Table 1 and an RGB image of the substances applied to a road is shown in Figure 3.

Our approach offers two key advantages: it enables system development and validation that are safe for both personnel and the environment, and it establishes a proof-of-concept for subsequent testing with actual hazardous substances.

Table 1. The substances used in our experiments.

Substance	Abbreviation
Potting Soil	So
Blue Nitrate Fertiliser	Fe
Potting Soil + Fertiliser	SoFe
Sand	Sa
Washing Powder	Wa
Ethanol	Et

The substances were selected based on two criteria. Firstly, they are readily accessible and categorised as safe for public road application. Secondly, each substance offers distinct spectroscopic properties valuable for experimental analysis: blue fertiliser provides a significant and unique spectral signature; washing powder delivers prominent white visibility with potential fluorescence from additives; sand mimics the appearance of washing powder without fluorescent properties; soil represents elements commonly found in standard road training data (such as residue from agricultural vehicles);

and ethanol was included to examine contrasts with rain-dampened road surfaces.

The soil and fertiliser were mixed together in different ratios (1p1 meaning 50/50 ratio, 1p4 meaning 20/80 ratio) to analyse differences in detection performance for different manifestations of spectral features (“blueness”) between them. Furthermore, we applied different amounts of the substances to yield different sample sizes.

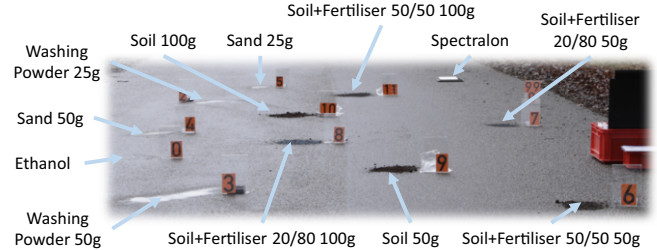


Figure 3. Test road with labelled substances which should be detected as AoIs.

3.4 Area of Interest Detection Pipeline

As aforementioned, identifying AoIs in multispectral imagery of roadways that contain unknown substances is framed as an anomaly detection problem. By training an autoencoder on normal images, reconstruction errors in parts of an image indicate anomalies, and thus, areas that contain unknown substances.

An overview is depicted in Figure 4. The particular processing steps are outlined in detail in the following.

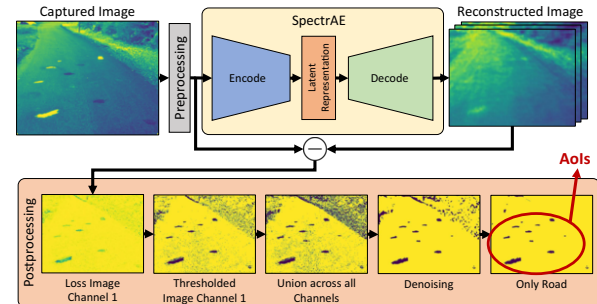


Figure 4. The full pipeline: A multispectral image is captured by the sensor array. In this case, the image contains hazardous substances. The autoencoder does not reconstruct them as accurately as the surrounding road, leading to localised losses. These are then post-processed to create AoIs for further analysis.

3.4.1 Preprocessing

The preprocessing transforms the raw multispectral images into a standardised format suitable for subsequent analysis. We extract the 9 spectral bands from each composite image using the demosaicing procedure described in Section 3.2. This transformation converts the original 1024×1280 image tensor into a $9 \times 339 \times 426$ tensor, where the 9th channel represents the greyscale data. To eliminate boundary artefacts that could potentially introduce noise into the analysis, we perform a centre cropping. The resulting tensor has the shape $9 \times 320 \times 416$. As a final preprocessing step, we normalise all pixel

values to the range $[0, 1]$, ensuring consistent input scaling across all spectral bands.

It is important to acknowledge that the camera manufacturer (SILIOS Technologies) recommends white reference correction and crosstalk correction to properly compensate for illumination variations and spectral leakage between adjacent filters. While these are important calibration procedures, our specific implementation of these corrections introduced unexpected artefacts that degraded model performance. This implementation challenge reflects limitations in our current data capturing pipeline. We continue to work on properly integrating these corrections into our processing pipeline for the future.

3.4.2 Autoencoder for Identifying Areas of Interest

SpectrAE is a convolutional autoencoder-based architecture designed for detection of unknown substances using multi-spectral imagery (Fig. 5). The model processes input from either the VIS or NIR camera, which capture complementary information about surface materials.

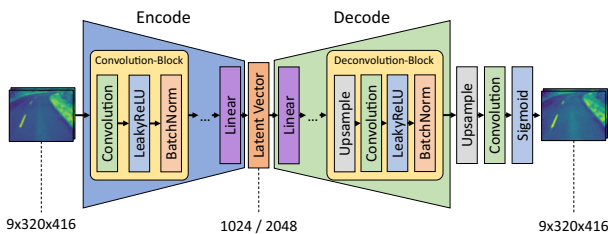


Figure 5. Architecture of SpectrAE: The shape of the tensor is provided at certain steps in the model to illustrate the feature dimensions.

SpectrAE implements reconstruction-based anomaly detection principles. During training, the model learns to accurately reproduce normal road surfaces by minimising reconstruction error on a dataset consisting exclusively of normal roads. This unsupervised approach enables the network to implicitly learn the characteristic spectral signatures and spatial patterns that define typical road surfaces without requiring labour-intensive manual labelling.

When deployed on test images containing anomalous substances, the decoder attempts to reconstruct what it expects—a normal road surface—rather than faithfully reproducing anomalous regions. This leads to a noticeable difference between the original input and the reconstruction precisely in areas containing potential hazards. The differences between actual and reconstructed image channels are referred to as loss images in the following.

The SpectrAE architecture (Fig. 5) follows established design principles for convolutional autoencoders [7]. The encoder comprises sequential 2D convolution blocks that progressively reduce spatial dimensions while increasing feature channels, culminating in a linear layer that transforms the processed features into a compact latent representation. The decoder then reconstructs the input dimensions through a linear layer followed by consecutive deconvolution blocks. A sigmoid activation function in the final layer constrains output values between 0 and 1, matching the normalised input range. All convolutions utilise a 3×3 kernel size with padding of 1 to preserve spatial information. The encoder employs a stride of 2 for downsampling, while the decoder uses a stride of 1 for precise reconstruction.

3.4.3 Postprocessing

After obtaining the reconstruction loss images (difference between input and reconstructed channels), AoIs are identified and refined in multiple stages (see bottom of Figure 4).

First, it has to be determined if the reconstruction error for particular pixels is out of the expected range, for a given spectral channel. To this end, we follow the common approach to take the 95th percentile of pixel losses across the validation set (containing only normal road surfaces). This method establishes channel-specific thresholds that account for the unique loss distribution characteristics of each spectral band that are then used to create binary anomaly maps per channel.

To maximise detection sensitivity, these channel-specific anomaly maps are combined by taking their union across all eight spectral bands (excluding the greyscale channel). This approach ensures that anomalous pixels in any spectral band are captured in the final detection output, even if they appear normal in other bands.

The binary anomaly mask after the union is quite noisy because of salt-and-pepper noise. For denoising, we apply a sequence of operations; first a morphological opening operation with a 2×2 -kernel, followed by a median blur with a 5×5 -kernel [18]. This combination strikes a good balance between removing salt-and-pepper noise and preserving edges.

As a final refinement step, we apply a pretrained image classification network to the greyscale channel to generate a binary road surface mask. This aligns with our problem assumptions outlined in Section 3.1 and ensures that detected anomalies are constrained to the road surface, further reducing false positives. The resulting refined AoIs provide localisation of potentially hazardous substances on the road surface.

4 Experiments

The following section outlines the experimental setup used to evaluate the proposed system.

4.1 Dataset

Our training dataset comprises 9,552 images (VIS and NIR) captured exclusively of normal road surfaces in rural areas. For the test set, we deliberately applied various substances to road surfaces and subsequently captured these modified areas. We manually annotated these images at the pixel level, labelling all present substances using the Labelme annotation tool [23]. The annotations were done by the authors and some of their students and visually checked for consistency across labellers by the authors. This process yielded 18 fully annotated images (9 VIS and 9 NIR), enabling quantitative evaluation of reconstruction quality across different substances. In total, the annotated images contained 31,417 pixels corresponding to the applied substances and 2,575,703 normal pixels.

4.2 Training

SpectrAE was trained for 100 epochs with a batch size of 256 on an NVIDIA QUADRO GV100 GPU with 32 GB of memory. Both the encoder and decoder comprise three layers each. Unless specified otherwise, all nine image channels were used as input. Since we do not fuse the VIS and NIR images, we have to train a separate model per image type. We set the latent dimension to either 1024 or 2048.

The model was optimised using the Mean Squared Error (MSE) loss function. A validation split of 0.1 was employed to evaluate generalisation performance throughout training.

For optimisation, we adopted the Adam optimiser [11] with a learning rate of 1×10^{-4} , $\beta = (0.9, 0.999)$, and a weight decay of 1×10^{-5} . To further enhance the training dynamics, we applied the ReduceLROnPlateau learning rate scheduling strategy, enabling adaptive adjustment of the learning rate and mitigating the risk of stagnation during model convergence.

The training resulted in four different models VIS₁₀₂₄, VIS₂₀₄₈, NIR₁₀₂₄ and NIR₂₀₄₈.

4.3 Reconstructed images

Figure 6 presents channel-wise loss images for a reconstructed test image. The three rows display, respectively, the input image, the reconstructed image, and the per-channel loss images, accompanied by a scaling reference. It is evident that different substances exhibit varying loss values across the channels, as anticipated, since each channel captures distinct spectral information.

This observation motivated our approach of computing the union across all channels after individually thresholding each one (see Sec. 3.4.3). This method compensates for the differing detection sensitivities present in each channel. In Figure 7, we display the result of this union for an example test image. Notably, certain AoIs visible in the union image are absent in the thresholded output of channel 1 alone, which would otherwise hinder detection performance. Although the union operation increases the level of noise, this drawback can subsequently be addressed by an appropriate denoising procedure.

4.4 Detection Performance

We evaluated SpectrAE’s detection performance using the labelled test images, reporting precision, recall, F_1 , F_2 , and Intersection over Union (IoU) for various SpectrAE configurations in Table 2. These metrics were computed on a pixel basis.

Table 2. Performance measures for different model architectures.

	Precision	Recall	F_1	F_2	IoU
VIS ₁₀₂₄	0.083	0.621	0.147	0.270	0.079
VIS ₁₀₂₄ RGB	0.185	0.574	0.280	0.404	0.163
VIS ₂₀₄₈	0.103	0.608	0.176	0.307	0.097
VIS ₂₀₄₈ RGB	0.210	0.566	0.306	0.422	0.181
NIR ₁₀₂₄	0.195	0.551	0.288	0.404	0.168
NIR ₂₀₄₈	0.213	0.559	0.309	0.422	0.183

For our application, recall is prioritised over precision, as it reflects the proportion of correctly detected anomalies—crucial since only identified AoIs can be further analysed. Sufficient detection ensures anomalies can be managed by the system. Nonetheless, precision also matters, since excessive false positives would make classification infeasible. Thus, we report F_2 (which weights recall higher than precision) and place greater emphasis on it than on F_1 .

To assess the necessity of multispectral information, we compared our results with an RGB approximation—using channels [0, 2, 6] of the VIS camera to best match conventional red, green, and blue wavelengths. The VIS model, trained on all 9 input channels, was used for a fair comparison. Interestingly, the RGB versions of the VIS models outperform those using all 8 spectral channels, with notably higher F_2 scores (0.404 and 0.422 vs. 0.270 and 0.307). This is unexpected,

given that the full model has access to the RGB channels plus additional spectral information. Closer inspection indicates that including all channels increases false positives, which is not mitigated by denoising, hence the lower performance.

VIS₁₀₂₄ achieves the highest recall but at the expense of precision. Both VIS₂₀₄₈RGB and NIR₂₀₄₈ perform comparably well. Whilst their results are encouraging, there is room for improvement. These models detect many AoIs (high recall), yet often fail to precisely localise them, as reflected by small IoU values.

To further assess model performance, we calculated substance-specific recall values, as shown in Table 3. Only recall is reported, since our pipeline produces binary masks and does not capture substance-specific false positives; a dedicated classification step would be needed for that. It is therefore essential to interpret these recall values alongside the overall precision results from Table 2, as high recall without adequate precision is of limited practical value. Table 3 reveals that most substances are best detected by the model with the lowest precision, namely VIS₁₀₂₄.

Ethanol (Et) consistently shows the lowest recall. This likely results from its rapid evaporation and spectral similarity to damp road surfaces, making it difficult for the network to distinguish. NIR-based models detect ethanol slightly better, possibly due to increased sensitivity to subtle spectral differences.

By contrast, washing powder (Wa) and sand (Sa) are detected most reliably in all models due to their strong contrast against the dark road. The two soil amounts (So) show varying recall, with the smaller amount detected better. Recall is higher for VIS models than NIR models, perhaps because NIR models encounter more soil along the roadside in the images and thus become less discriminative.

For the soil and fertiliser mixture (SoFe 1p1), VIS models significantly outperform NIR models. For the 20/80 mixture (SoFe 1p4), only VIS models using all 8 channels detect the substance effectively, and again, the smaller quantity is easier to detect. Detection declines with increased fertiliser content, probably due to the fertiliser’s blue tint reducing contrast.

4.5 Component Analysis

To systematically evaluate our modelling choices, we conducted a comprehensive component analysis of our pipeline (Table 4) using our best-performing model, NIR₂₀₄₈, trained on all 9 input channels to ensure a fair comparison.

Instead of taking the union of the channel-specific anomaly maps (Sec. 3.4.3) to decide if a pixel is an anomaly, we also examined alternative schemes, namely intersection and errors out of range in the majority of the channels (majority vote), the impact of removing denoising, and the performance of individual channels. Using the intersection of anomaly maps of the channels yields highest precision (0.536), but at the cost of recall (0.204), as fewer pixels are detected as anomalies. The majority vote underperforms compared to the union on F_2 .

Removing denoising maximises recall (0.62), but sharply reduces precision (0.077) due to increased noise, indicating that denoising—while imperfect—is necessary, though further improvements are needed, as some true positives are lost, likely near AoI edges.

Comparisons of individual channels show that all but channel 1 perform worse than the full model. Notably, channel 1 slightly outperforms the full model in F_2 (0.427 vs. 0.422), with higher precision but lower recall, meaning it produces fewer false positives but misses more substances. This reflects that the union of anomaly maps across channels adds false positives in the full model. Shorter

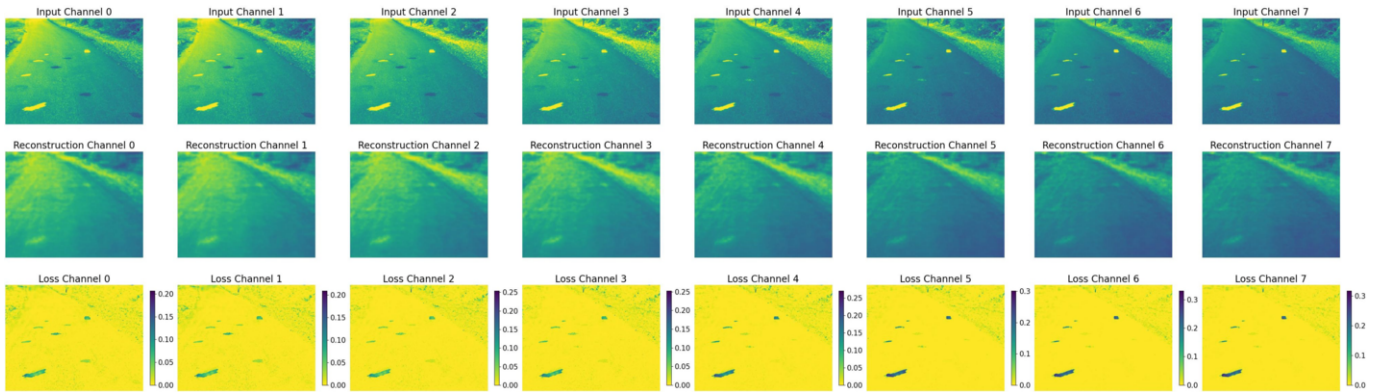


Figure 6. Comparison of the reconstruction quality for all 8 spectral channels of an NIR image: One can see that different channels detect different substances, e.g. channel 1 detects substances that are not visible in the loss image of channel 7.

Table 3. Recall values by class. The abbreviations are explained in Table 1.

	Precision	SoFe 1p1 100	SoFe 1p1 50	SoFe 1p4 100	SoFe 1p4 50	So 100	So 50	Sa 20	Sa 50	Wa 25	Wa 50	Et
VIS ₁₀₂₄	0.083	0.928	0.959	0.499	0.646	0.703	0.919	0.830	0.863	0.982	0.997	0.117
VIS ₁₀₂₄ RGB	0.185	0.898	0.919	0.184	0.393	0.692	0.882	0.788	0.783	0.955	0.981	0.105
VIS ₂₀₄₈	0.103	0.911	0.932	0.466	0.526	0.699	0.898	0.828	0.848	0.974	0.993	0.116
VIS ₂₀₄₈ RGB	0.210	0.884	0.886	0.188	0.354	0.691	0.865	0.783	0.759	0.954	0.983	0.101
NIR ₁₀₂₄	0.195	0.489	0.489	0.466	0.385	0.517	0.751	0.860	0.717	0.941	0.962	0.144
NIR ₂₀₄₈	0.213	0.468	0.523	0.446	0.424	0.581	0.712	0.875	0.633	0.941	0.955	0.162

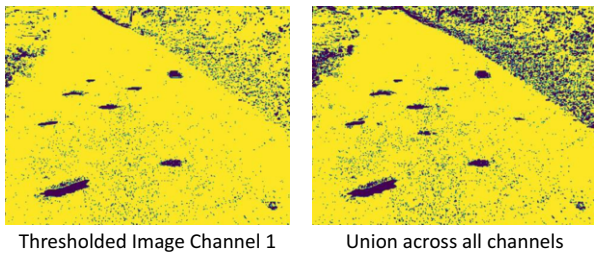


Figure 7. The importance of taking the union across all channels: Several AoIs that are not visible in the thresholded image of channel 1 become apparent when the union across all channels is considered.

wavelengths (early channels) generally outperform longer ones (e.g. channel 1 vs. channel 6), possibly due to incomplete calibration (see Sec. 3.4.1) or lower information content in the later wavelengths. Finally, the full model surpasses the greyscale channel, underlining the value of using multiple spectra for improved detection.

4.6 Latent dimension

We investigated whether the autoencoder’s latent representation (Fig. 5) could distinguish between images with and without hazardous substances. Such separation would enable rapid pre-detection of normal images before decoder processing, thereby improving computational efficiency and reducing false positive AoIs in normal road scenarios.

Table 4. Component analysis results based on our best-performing model.

	Precision	Recall	F ₁	F ₂	IoU
Intersection	0.536	0.204	0.296	0.233	0.173
Majority vote w/o denoising	0.361	0.620	0.352	0.347	0.213
Channel 0	0.340	0.429	0.380	0.409	0.234
Channel 1	0.361	0.448	0.400	0.427	0.250
Channel 2	0.408	0.375	0.391	0.381	0.243
Channel 3	0.390	0.330	0.358	0.340	0.218
Channel 4	0.374	0.277	0.318	0.292	0.189
Channel 5	0.372	0.278	0.318	0.293	0.189
Channel 6	0.363	0.286	0.320	0.298	0.191
Channel 7	0.354	0.294	0.322	0.303	0.192
Greyscale	0.364	0.377	0.370	0.374	0.227
NIR ₂₀₄₈	0.213	0.559	0.309	0.422	0.183

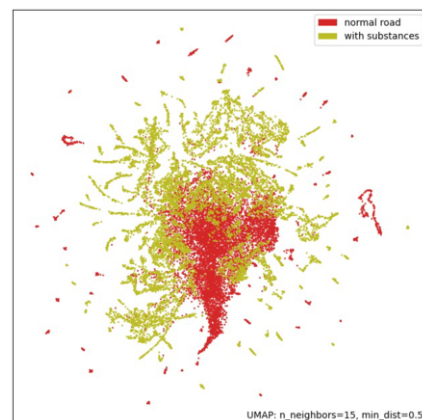


Figure 8. 2D UMAP embedding of the dataset splits: The normal roads remain indistinguishable from roads with applied substances.

To visualise the high-dimensional latent space, we computed a 2D UMAP [17] embedding from the 2048-dimensional latent vectors across all dataset partitions (Fig. 8), using the NIR₂₀₄₈ model. The UMAP parameters were configured with 15 neighbours, Euclidean distance metric, and a minimum distance of 0.5 to balance local and global structure preservation.

Our analysis reveals that despite hazardous substances appearing exclusively in the test set, the test images remain indistinguishable from normal road images in the latent space. This lack of separation likely stems from several factors, such as the limited pixel coverage of hazardous substances relative to the entire image creating minimal signal in the high-dimensional latent space. Additionally, environmental variations—particularly roadside vegetation—introduce substantial image variability that overshadows the subtle differences from hazardous substances.

5 Discussion

In the following section key aspects of the proposed system are critically discussed.

5.1 Camera and Image-Related Challenges

The image acquisition process presents several challenges that may have influenced detection performance. Notably, white reference and crosstalk correction were not incorporated in our current setup, despite being reported by the manufacturer as essential for obtaining correct pixel values. This limitation became more pronounced at longer wavelengths, where we observed increased issues during our tests. Consequently, this may have negatively impacted detection accuracy. It is crucial to address this issue for future classification tasks. Furthermore, the cameras' auto-exposure algorithm computes the exposure based on the central lower part of the image. As a result, the upper region of the images often exhibits noticeable artefacts. To mitigate this, one could crop out the affected upper area and design the analysis pipeline to focus on the region that aligns with the auto-exposure setting. Doing so would help maintain the validity of the extracted pixel values. Finally, the overall number of labelled images in the dataset is relatively small. Increasing the volume of annotated data would provide a clearer understanding of the system's performance and improve model robustness for practical deployments.

5.2 Reconstructed Image Quality

The current reconstruction quality achieved by our models is already promising, although the architectures used are relatively simple autoencoders. There is considerable potential to enhance performance by adopting more sophisticated encoders. Future work should explore the integration of pretrained encoders, such as ResNet [8] and vision transformers [22, 14, 4, 28], to better capture complex image features critical to substance detection and classification.

Another avenue for improvement concerns the prediction strategy. Instead of processing entire images at once, dividing input images into smaller patches could prove advantageous. This approach would lead to a proportionally bigger area of the input being AoIs, potentially leading to clearer separation in the latent space and improved reconstruction of critical regions.

Currently, reconstruction loss is measured using the MSE metric. Investigating alternative loss functions, such as the Structural Similarity Index Measure (SSIM) [26], could yield further improvements.

SSIM, in particular, is known for its ability to capture perceptual differences in image quality more effectively than MSE, which may lead to reconstructions that are visually closer to the original images.

Exploring these refinements has the potential to both improve reconstruction fidelity and enhance the applicability of our models to real-world scenarios.

5.3 Detection Performance Analysis

A key question arising from our results is whether using all available spectral channels in combination truly yields optimal detection performance. There is evidence from our ablation study that certain individual channels outperform the combined model, suggesting that some channels may contribute disproportionately more relevant information, whilst others may introduce additional noise. Selecting a subset of the most informative channels could therefore enhance detection robustness. This warrants further investigation and careful channel selection in future work. A deeper analysis of substance classification performance is also required. Rather than focusing solely on recall values per class (Tab. 3) it would be valuable to conduct a more general evaluation. Some specific substances, such as ethanol, were not detected reliably; this is likely due to its rapid diffusion into the surrounding environment. It would be beneficial to explore detection with additional substances such as diesel or tonic water (containing quinine), to assess whether similar challenges are encountered, and to better understand the generalisability of the method. It is also important to consider whether the chosen substances are suitably representative for comparative purposes. Until now, the solid substances tested have been relatively easy to discern visually. Future experiments could include substances that are challenging to discriminate from the road surface by eye, in order to demonstrate the practical advantage of automated spectral detection in contrast to a human observer. Finally, our detection results have not yet been compared to established methods, such as the RX algorithm [19] or other approaches [10]. Benchmarking against these alternatives would provide important context for assessing the strengths and limitations of our methodology.

6 Conclusion

This work represents an initial step towards substance detection using spectral imaging and identifies several avenues for further improvement. One promising direction is the integration of outputs from multiple cameras and the fusion of their data to enrich the dataset. The hardware warrants enhancement through the utilisation of advanced cameras and improved post-processing techniques to ensure data reliability. The test dataset should be expanded both by increasing the number of test images and by including images without substances to evaluate whether the model correctly identifies normal roads. It is necessary to assess classification performance to determine whether current camera systems can reliably distinguish between relevant substances within AoIs. An open question remains whether spectral features alone suffice for accurate substance classification, or whether techniques such as LIF will be required to achieve robust results.

A current limitation of this work is its focus on controlled road scenarios, which constrains the generalisability of the results. Expanding the scope to encompass more diverse and realistic conditions will be crucial for translation into practical, real-life deployments with meaningful societal impact.

Acknowledgements

We want to thank Thomas Schlagenhauer and Lara-Cathrin Walprecht for the driving tests and Anna Zethmeyer for the help with quality management. Furthermore, we thank Cansu Bayez and Franziska Mammes for their help in labelling the dataset.

References

- [1] M. Abdulsalam, U. Zahidi, B. Hurst, S. Pearson, G. Cielniak, and J. Brown. Unsupervised tomato split anomaly detection using hyperspectral imaging and variational autoencoders. In *Computer Vision – ECCV 2024 Workshops: Milan, Italy, September 29–October 4, 2024, Proceedings, Part III*, page 101–114, Berlin, Heidelberg, 2025. Springer-Verlag. ISBN 978-3-031-91834-6.
- [2] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2019) - Volume 5: VISAPP*, pages 372–380. INSTICC, SciTePress, 2019. ISBN 978-989-758-354-4. doi: 10.5220/0007364503720380.
- [3] S. Chen, X. Li, and Y. Yan. Hyperspectral anomaly detection with autoencoder and independent target. *Remote Sensing*, 15(22), 2023. ISSN 2072-4292. doi: 10.3390/rs15225266. URL <https://www.mdpi.com/2072-4292/15/22/5266>.
- [4] B. Choi and J. Jeong. Viv-ano: Anomaly detection and localization combining vision transformer and variational autoencoder in the manufacturing process. *Electronics*, 11(15), 2022. ISSN 2079-9292. URL <https://www.mdpi.com/2079-9292/11/15/2306>.
- [5] J. Chow, Z. Su, J. Wu, P. Tan, X. Mao, and Y. Wang. Anomaly detection of defects on concrete structures with the convolutional autoencoder. *Advanced Engineering Informatics*, 45:101105, 2020. ISSN 1474-0346. doi: <https://doi.org/10.1016/j.aei.2020.101105>. URL <https://www.sciencedirect.com/science/article/pii/S1474034620300744>.
- [6] J. Feng and L. Zhang. Hyperspectral anomaly detection based on autoencoder and spatial morphology extraction. *Journal of Applied Remote Sensing*, 15(3):038507, 2021. doi: 10.1117/1.JRS.15.038507. URL <https://doi.org/10.1117/1.JRS.15.038507>.
- [7] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [8] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [9] M. V. Hikuwai, N. Patorniti, A. S. Vieira, G. Frangioudakis Khatib, and R. A. Stewart. Artificial intelligence for the detection of asbestos cement roofing: An investigation of multi-spectral satellite imagery and high-resolution aerial imagery. *Sustainability*, 15(5):4276, 2023. ISSN 2071-1050. doi: 10.3390/su15054276.
- [10] T. Hupel and P. Stütz. Adopting hyperspectral anomaly detection for near real-time camouflage detection in multispectral imagery. *Remote Sensing*, 14(15):3755, 2022.
- [11] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [12] C. Kölbl, M. Diedrich, E. Ellingen, D. Weigl, and F. Duschek. Laserspektroskopische Ferndetektion zur Erkennung von Gefahrstoffen auf Kleindrohnen. In *DWT Tagung Unbemannte Systeme*, April 2023. URL <https://elib.dlr.de/194849/>.
- [13] W. Lai, M. Zhou, F. Hu, K. Bian, and H. Song. A study of multispectral technology and two-dimension autoencoder for coal and gangue recognition. *IEEE Access*, 8:61834–61843, 2020.
- [14] Y. Lee and P. Kang. Anovit: Unsupervised anomaly detection and localization with vision transformer-based encoder-decoder. *IEEE Access*, 10:46717–46724, 2022.
- [15] J. Lei, S. Fang, W. Xie, Y. Li, and C.-I. Chang. Discriminative reconstruction for hyperspectral anomaly detection with spectral learning. *IEEE Transactions on Geoscience and Remote Sensing*, 58(10):7406–7417, 2020.
- [16] S. Matteoli, M. Diani, and G. Corsini. A tutorial overview of anomaly detection in hyperspectral images. *IEEE Aerospace and Electronic Systems Magazine*, 25(7):5–28, 2010.
- [17] L. McInnes, J. Healy, N. Saul, and L. Grossberger. Umap: Uniform manifold approximation and projection. *The Journal of Open Source Software*, 3(29):861, 2018.
- [18] W. K. Pratt. *Digital image processing: PIKS Scientific inside*, volume 4. Wiley Online Library, 2007.
- [19] I. S. Reed and X. Yu. 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, 1990.
- [20] D. W. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker. Anomaly detection from hyperspectral imagery. *IEEE signal processing magazine*, 19(1):58–69, 2002.
- [21] H. T. Tran and D. Hogg. Anomaly detection using a convolutional winner-take-all autoencoder. In *Proceedings of the British machine vision conference*. British machine vision association, 2017.
- [22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [23] K. Wada. Labelme: Image polygonal annotation with python. 2025. doi: 10.5281/zenodo.5711226. URL <https://github.com/wkentaro/labelme>.
- [24] R. Wang, F. Nie, Z. Wang, F. He, and X. Li. Multiple features and isolation forest-based fast anomaly detector for hyperspectral imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 58(9):6664–6676, 2020.
- [25] S. Wang, X. Wang, L. Zhang, and Y. Zhong. Auto-ad: Autonomous hyperspectral anomaly detection network based on fully convolutional autoencoder. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–14, 2021.
- [26] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [27] P. Xiang, S. Ali, S. K. Jung, and H. Zhou. Hyperspectral anomaly detection with guided autoencoder. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–18, 2022.
- [28] J. Zhang, X. Chen, Y. Wang, C. Wang, Y. Liu, X. Li, M.-H. Yang, and D. Tao. Exploring plain vit reconstruction for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2312.07495*, 2023.
- [29] L. Zhang and B. Cheng. A stacked autoencoders-based adaptive subspace model for hyperspectral anomaly detection. *Infrared Physics & Technology*, 96:52–60, 2019.