

Intelligent Hybrid Mobile Robotic Landmine Detection System*

Victor Sineglazov^{1,†}, Kyrylo Lesohorskyi^{2,*,†}

¹ National University of Ukraine «Kyiv Aviation Institute», ave. Lubomir Husar, 1, 03058, Kyiv, Ukraine

² National Technical University of Ukraine “Ihor Sikorsky Kyiv Polytechnic Institute,” Ave. Beresteysky, 37, Kyiv, Ukraine

Abstract

This paper considers a hybrid mobile robotic system for the problem of landmine detection. The system consists of two intelligent robotic agents. The first agent is a highly mobile aerial-based detector with a hyperspectral / multispectral camera and is designed to identify all areas of landmine installation. The second intelligent agent is a ground-based robot with an infrared camera, which has a lower sensitivity threshold and is used to further validate the mines identified by the first intelligent agent. Semi-supervised learning with spectral-spatial consistency is used to train CNN-based feature extraction and classification pipeline. The proposed technique allows to boost labeled sample size by ~3 times and achieves high recall (1.0) and moderate precision (0.694). The verification step improves precision to 0.93, reducing the number of false positives.

Keywords

Landmine detection, hybrid robotic system, semi-supervised learning, hyperspectral imagery

1. Introduction

In many countries, mines pose a serious threat to life and cause economic problems. Mines are dangerous because their location is unknown and they are often difficult to detect. The development of new demining technologies is difficult due to the wide variety of terrains and environmental conditions in which mines are laid, as well as the wide variety of mines. Currently, detecting and clearing mines requires special knowledge and special equipment.

The conduct of active hostilities on the territory of our state led to the appearance of significant territories contaminated with explosive objects, which poses a real danger to people's lives and health, prevents them from ensuring their livelihoods and restoring economic activity, has a negative impact on ecosystems, etc. Today, Ukraine is the most polluted place in the world. According to estimates, about a third of its territory - an area the size of Florida - contains WB.

Prompt, accurate and automated detection and determination of explosive objects is becoming an extremely urgent task today. At the same time, the methods and conditions of planting explosive objects, in particular their special concealment, scattering as a way of setting up and scattering remnants as a result of explosions, as well as "noisiness" of the battlefield territories and significant limitations (primarily of a technical nature) of detection means, significantly complicate the task of detecting and identifying explosive objects. by conventional computer vision methods. This necessitates the development of new methods for determining such objects, based on hybrid approaches. The process of detecting mines is complex, dangerous and expensive. The main problem is not to remove mines, but to accurately determine the location of the mine. This entire operation is still performed manually, which leads to the fact that the pace of demining is very unsatisfactory. Since 20 times more mines are installed for each remote mine [3].

*2nd International Conference on Smart Automation & Robotics for Future Industry, April 03-05, 2025, Lviv, Ukraine

^{1*} Corresponding author.

[†] These authors contributed equally.

✉ svm@nau.edu.ua (V. Sineglazov); lesogor.kirill@gmail.com (K. Lesohorskyi);

0000-0002-3297-9060 (V. Sineglazov); 0000-0003-2773-7398 (K. Lesohorskyi);



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Therefore, a more appropriate approach based on the latest technologies is critical, minimizing risk while increasing the speed of demining and the accuracy of the procedure. Therefore, the development of hybrid robotic systems that include unmanned air and ground vehicles and can carry sensors with minimal interaction with human operators is of great importance. [6].

This is achieved by dividing the overall system into two subsystems: sensor technologies and a robotic device.

There are many problems in mine detection. The first is that changes in weather factors have led to the disappearance of underground mine spaces. The second is the ability to stop the effects of mines without seeing underground. The third is that detection requires not only the presence of a mine, but also the robot must mark the location of the mine with an accuracy of 5 cm.

Work is needed to merge mine detection technologies to improve their performance, as each approach produces good results in limited conditions. Due to the above limitations, a multi-sensor system based on signal fusion and algorithms should be developed. Rather than focusing on individual technologies operating in isolation, mine detection research and development should emphasize design from first principles and subsequent development of an integrated multi-sensor system that overcomes the limitations of any single sensor technology. Combining different types of sensors will certainly achieve better detection results [1].

2. Robot-based landmine removal

The ability to detect mines in a surface minefield using autonomous robots is becoming increasingly popular as it reduces the danger and cost of manual detection [19]. The robots search for mines with such low pressure that mine explosions do not go off. To effectively cover all mined areas, robots must adapt to accelerated reconnaissance to improve efficiency, especially if there is any surveillance team. The use of robots to detect landmines provides an ideal sensor for robots due to its low cost, wide availability, high data content and speed of information. Adequate clearance rates can only be achieved using new technologies such as improved sensors, efficient manipulators and mobile robots [16]. Estimating the position of buried landmines using data from landmine detection sensors is important in selective work. One of the problems with current mine detection robots is that they are quite large in structure and very expensive. Although several inexpensive mine detection robots have also been developed, most of them use simple algorithms so that they can only work in simple, unobstructed environments [11]. The development of lightweight, low-cost, semi-autonomous robots operating in conjunction with a monitoring station (personal mine scouts) is a well-studied approach [17]. The robots search for mines at such low pressure that mine explosions do not go off. Multi-robot systems for area reduction form the next step in the search for landmines. Some research has been carried out on a multi-agent based architecture responsible for coordinating advanced stochastic terrain analysis. It includes the reactive obstacle avoidance technique and the development of mission control software to plan, configure and control operations. The system uses walking, wheeled and aerial robots. Finally, this study describes a payload sensor system using Fourier analysis as a mechanism for efficient mine detection.

Three main facts made it possible to gradually increase the efficiency of surveys using UAVs [13,14]:

- The UAV industry (eg DJI) has created very advanced systems that enable computer-aided planning and automated aerial detection missions with multiple sensors.
- The sensor industry has provided powerful devices to match UAVs.

- The software industry has provided tools to process the records collected by UAVs, producing results of the highest quality.

The first UAV for humanitarian demining appeared in the EU ARC project. The use of UAVs with visible color sensors for humanitarian mine clearance has increased over the past 5–10 years. [4,5,9–12].

In the context of a high level of false alarms in an intelligent hybrid mobile robotic mine detection system, it is proposed to use two information-connected robots.

The idea of obstacle avoidance for an autonomous mine detection robot is based on the use of continuous mapping and localization techniques (simultaneous localization and mapping – SLAM). This allows the robot's position to be localized while the map is updated using Gmapping [17].

The purpose of this is to adjust the parameters so that the robot can navigate the new map and reach the desired location, avoiding potential obstacles along the way. The path planner gives the robot a path, which is a new path that allows the robot to avoid obstacles along the way. The local planner sends commands that allow the robot to follow its path. This is done by estimating the robot's position using data obtained from a laser scanner [18].

3. Landmine detection technologies

Today, the main means of technical inspection in mine countermeasures include:

1. Metal detectors that work on the basis of the principle of electromagnetic induction.
2. Geo-radar (Ground Penetrating Radar, GPR), which use high-frequency electromagnetic waves to scan subsurface layers of soil, concrete, or other materials.
3. Explosive vapor detectors ("electronic noses") that detect molecules or microparticles of explosive substances in the air.
4. Mine detectors based on acoustic, ultrasonic or seismic methods.
5. Non-linear mine detectors (Non-Linear Junction Detector, NLJD) are devices for detecting non-linear electronic components that are commonly used in various electronic devices.
5. X-ray systems that are not used for wide areas and have restrictions on the location of the object in the viewing area.
6. Thermal imagers, which have less effect in the case of equalized temperature (for example, at night or during long-term cover of explosives).

However, traditional methods used in mine countermeasures have significant limitations, which depend not only on the physical principles of action and the technical and technological levels of development of these methods, but above all on the specific tasks and conditions of this activity.

Innovative methods of detecting explosive objects include, first of all, the methods of analyzing infrared and hyperspectral images (Hyperspectral Imaging, HSI). Despite the maturity of sensors, signal processing algorithms remain underdeveloped and not related to physical phenomena. Thermal signatures are currently not sufficiently studied, and there is no comprehensive prognostic model" [20]

Practically all studies of methods of detecting mines and explosive objects mentioned here, both traditional and innovative, agreed with the thesis that for their effective determination, it is necessary to apply the combination and fusion of data of different types.

Research suggests using both different types of sensors and different methods of data fusion. The hardware combination is based on the fact that multisensor systems combine technologies with various sources of false alarms, so they can significantly reduce the frequency of false alarms. The most powerful methods of processing infrared and hyperspectral images are methods of artificial intelligence.

4. Objects and features

The measured data in hyperspectral images can be visualized as a data cube. A hyperspectral cube is a three-dimensional array of data that represents a set of spectral images of an object captured in various narrow spectral bands in the electromagnetic spectrum. The hyperspectral cube consists of two-dimensional spatial information (x , y), which reflects the location of pixels in the image, and the third dimension - the spectral axis (λ), which corresponds to the intensity of light or signal at each wavelength. Each slice of the data cube contains an image of the scene at a specific wavelength. Each pixel is associated with a vector of spectral responses, otherwise known as a spectral signature. The signal intensity (which is recorded for each pixel and each spectral channel) is the main parameter that reflects the amount of electromagnetic radiation recorded by the sensor. The hyperspectral cube is thus defined by spatial and spectral structures. Spatial structure is characterized by:

a) resolution (spatial resolution), which determines the detailing of spatial information, i.e. high spatial resolution makes it possible to recognize small objects in the image;

b) size in space: dimensionality along the axes x and y corresponds to the number of pixels that cover the area of the object or scene.

The spectral structure of the hyperspectral angle is determined by:

a) the number of spectral channels (bands), i.e. the number of narrow spectral bands in which the signal is recorded - usually from several tens to hundreds of channels;

b) spectral resolution, which determines the width of each spectral channel (for example 1–10 nm); higher spectral resolution makes it possible to better distinguish materials;

c) wavelengths that cover a certain range of the electromagnetic spectrum.

The conditionality of the data, in particular, a large number of spectral bands, determines the high computational complexity of the hyperspectral cube data processing algorithms. An important characteristic of the hyperspectral cube is the signal-to-noise ratio (SNR), which determines the quality of spectral information: a high SNR value provides a more accurate determination of material characteristics. Some hyperspectral cubes include a time component, which adds another measurement axis. This is used to monitor dynamic processes (such as changes in vegetation) or highly dynamic processes (targeting). The data of hyperspectral cubes are highly informative, because they contain detailed information about the physical, chemical, biological, etc. properties of objects. Another property of hyperspectral cube data is the spectral correlation of channels, which should be taken into account during intelligent image processing. Signal intensity can be expressed in relative units (reflectance, emissivity, etc.) or calibrated physical units. Calibrated units take into account the physical properties of the scene, the sensor, and the environment in which the measurement is made. They make it possible to obtain accurate data about the physical parameters of the signal, for example, energy, flow or radiation intensity. Calibration involves taking into account the spectral sensitivity of the sensor, atmospheric conditions (scattering, absorption), scene geometry (angle of light incidence/reflection), power of the light source. It should be taken into account that calibration requires additional time, equipment and resources.

As of today, the main features that are promising for use in the remote detection of mines and explosive devices using the methods of intelligent processing of hyperspectral images can be considered:

1) actual mines and explosive objects as physical objects;

2) spectral signatures of explosives and associated substances;

3) disturbance of land cover and vegetation;

4) marking of dangerous objects and zones (in conventional cases);

5) unmasking signs on the terrain (in case of concealment).

The main task of using hyperspectral analysis methods to solve the tasks of demining territories and disarming explosive devices can be the detection and identification of substances used in explosive devices. It is the spectral signatures (signatures) of the specified substances that will determine the limits of the wavelength ranges of electromagnetic radiation, which should be investigated using hyperspectral analysis, and therefore, the technical requirements for equipment and, ultimately, the methods of intelligent analysis of hyperspectral images. However, methods of detection and identification of substances in demining tasks have certain limitations.

5. Problem Statement

In this work, the task of detecting mines will be considered as a segmentation task - pixels that correspond to mines and other objects of interest must be selected on the target images. More formally, an image I represented as a tensor $H \times W \times C$, where H is the height of the matrix, W is the matrix width, C is the number of channels. Each element of the matrix x represents an image pixel and contains a single rational value $x \in R$, which corresponds to the "brightness" of the surface in this pixel. The work solves the task of constructing a segmentation function that transforms the input image I to the segmentation map I' of size $H \times W$ where each element $y \in C$, $C = \{\emptyset, 1, 2, 3, \dots, c\}$, defined as an empty space or an explosive ordnance.

When evaluating the quality of an intelligent system, it is important to use realistic standards. The UN standard for demining landmines requires a 99.6% explosive ordnance detection rate for humanitarian demining. However, this standard does not establish the level of false positives. Such triggers are not dangerous for the life and health of the personnel performing demining, but they significantly slow down and make the process more expensive. To overcome this shortcoming, the use of a hybrid system of two autonomous devices is proposed. The first device is mobile and allows for a quick processing of a large area, and the second.

Since complex high-dimensional images are used to solve the segmentation problem, the problem of a limited data set arises. Collecting data using one sensor is labor intensive, but adding a second, different type of sensor increases the complexity of the process even more. Semi-supervised learning is used to solve this limitation. Formally, this process is described as the training of an approximator $f(x, \theta)$, where $x \in X$ is the input space, $y \in Y$ is the output space, θ are approximator's parameters, derived through a training process. The learning function is used to learn the weights $T(f, L, U) = \theta$, where T is the training function, f - is the approximator, L - is the labeled dataset, $L = (x_1, y_1), \dots, (x_n, y_n)$, $x \in X$, $y \in Y$, U is the unlabeled dataset, $U = x_{(n+1)}, \dots, x_m$, $x \in X$. It should be noted that the approximator obtained by means of semi-supervised learning does not differ from the approximator obtained using classical supervised learning at the inference stage.

When evaluating the effectiveness of the proposed solution, it is important to choose the right metrics, as it is important to correctly classify the mine, and to avoid errors of the first kind (false negatives).

The following metrics will be used in this work:

Precision (Mine / No Mine) – the ratio of all true results to the total number of true and false-true

$$\text{results: } Precision = \frac{TP}{(TP + FP)} ;$$

Recall (Mine / No Mine) – the ratio of all true results to the total number of true and false negative results. When detecting mines, it is very important not to make mistakes of the first kind, which

$$\text{makes recall one of the key metrics: } Recall = \frac{TP}{(TP + FN)}$$

Rand index (Per Class) – also known as classifier accuracy. This metric evaluates the ratio of true positive and true negative predictions to all predictions: $Rand = \frac{TP + TN}{(TP + TN + FP + FN)}$

Intersection over Union (IoU) – is a classic accuracy indicator that is often used to assess the quality of a solved segmentation problem. The IoU calculates the ratio of the intersection of the expected class label to the total sum of the prediction area and the actual label.

$$IoU = \frac{TP}{(TP + FP + FN)}$$

6. Method

6.1. General description of the approach

The proposed approach is based on the use of two devices for remote survey of the mined area and detection of regions that contain mines. The first device is a highly mobile (quadrocopter), the task of which is to detect areas with a high probability of finding explosive objects. For this, a hyperspectral or multispectral camera is used, capable of detecting mines on the surface or in the ground at a depth of up to 10 centimeters. At the same time, the intelligent system is highly sensitive and allows a high number of errors of the second kind.

To compensate for this, a second, less mobile, ground-based device is used. It is equipped with a more sensitive sensor (lidar, ground-penetrating radar, magnetometers). At the same time, its intelligent system is less sensitive, which significantly reduces the level of type II errors without creating additional risk to personnel. This approach has following advantages:

1. This removes restrictions on the use of heavy and bulky sensors that cannot be installed on a quadrocopter;
2. This makes it possible to speed up the collection of data by a ground robot by optimal route planning and scanning of areas with a high probability of exposure to explosive objects, without visiting areas that were confirmed to not have any landmines;
3. This enables usage of a combination of several sensors. The specific types of sensors are chosen depending on the task, but at the post-processing stage, the sensor fusion technique can be used to obtain an even richer set of data.

The general scheme of the system is shown in Figure 1.

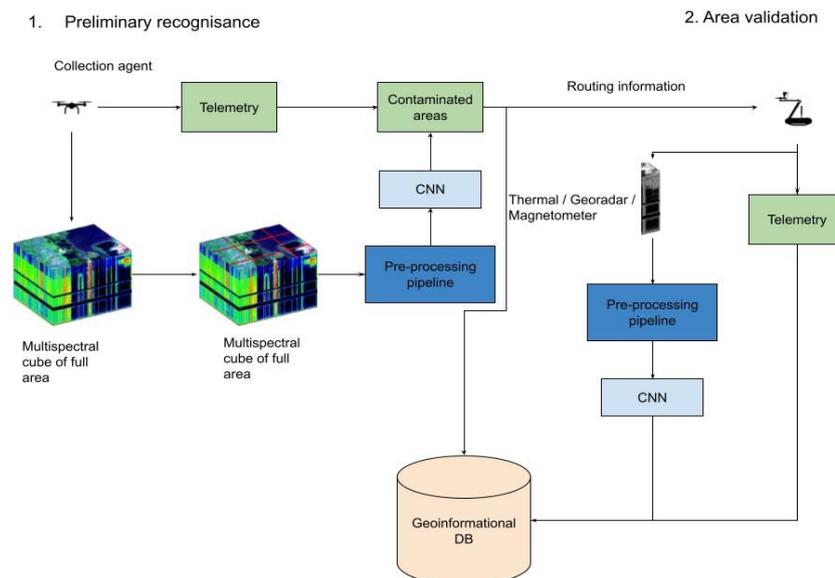


Figure 1: The diagram of the proposed approach

6.2. Data Processing

A popular trend in hyperspectral data processing is a comprehensive data processing pipeline that denoises, normalizes, filters, and (if necessary) reduces the dimensionality of the input data. This work uses a hybrid approach with two intelligent systems that control the level of sensitivity individually, a lightweight reprocessing pipeline is used. The pipeline is designed to create a high-quality contrast with minimal blurring or artifact introduction into the original image. The proposed preprocessing consists of local-global contrast stretching via normalization. When normalizing, a grayscale representation of pixels is used. After calculating the corresponding coefficients, each channel is scaled by multiplying the value by the corresponding coefficient.

In the case when the elements of the scene are illuminated unevenly, this can lead to different local contrast in the same objects. Since contrast strength is an important feature, it must be normalized. For this, local normalization based on convolution is used:

$$I_n(x, y) = newMax * \frac{(I(x, y) - min_I(x, y))}{max_I(x, y) - min_I(x, y)}, \quad (1)$$

where $newMax$ is the maximum value after the normalization, $I(x, y)$ is pixel value x, y in the original image, $min_I(x, y)$ is the minimum value for the convolution kernel in pixels x, y , $max_I(x, y)$ is the maximum value for the convolution kernel in a pixel x, y .

For local convolution, it is recommended to use a medium-sized window (this work uses a kernel of 25×25), however, specific settings depend on the data set.

Local normalization is followed by global normalization (within a batch during training or a window during inference). Global normalization is implemented as linear normalization, which allows to reduce the density of the space of the input distribution:

$$I_N = (I - Min) \frac{newMax - newMin}{Max - Min} + newMin, \quad (2)$$

where Min is the minimum brightness value in the original image, Max is the maximum brightness value in the original image, $newMax$ is the new maximum value in the image, $newMin$ is the new minimum value in the image. Linear normalization uses standard parameters $newMax = 255$, $newMin = 0$.

6.3. Convolutional Neural Networks for feature extraction and segmentation

Artificial convolutional neural networks are used in the work to extract features and perform segmentation. This type of neural networks makes it possible to flexibly process features from high-dimensional data when solving a classification problem. This makes it possible to use the same architecture for segmentation and feature extraction from a wide range of sensors that can be reduced to a tensor format.

The use of a homogeneous architecture significantly simplifies the learning process and also expands the number of available types of sensors, as it allows for unified processing of both low-dimensional and high-dimensional data.

The classic Unet architecture with residual connections was chosen as the architecture of the segmenter. The architecture of the neural network is shown in Figure 2.

The proposed architecture consists of three components:

1. Input adapter – this block consists of several 3D convolution layers and brings the input layer to a fixed dimension. The main task of this block is, first of all, spectral compression,

which highlights features while reducing the resolution of the input, so the input data is processed incrementally by granular kernels.

2. Feature detection path – this block consists of the classical U-net architecture and contains convolutional and sweeping paths. Each of these paths consists of three blocks of convolution (or deconvolution) and feature detection. There are residual connections between the corresponding blocks of each pathway, which prevents gradient attenuation and stabilizes learning. The convolution block consists of a maximum pooling block and 2 consecutive convolution operations. A deconvolution block consists of two consecutive convolution operations and one sweep operation.
3. Segmentation Adapter – expands the resulting feature map to the target size, producing a segmentation map that fits the size of the input. This layer also uses three-dimensional convolutions to combine features from different channels into one, forming at the output a segmentation map with the size $H \times W$.

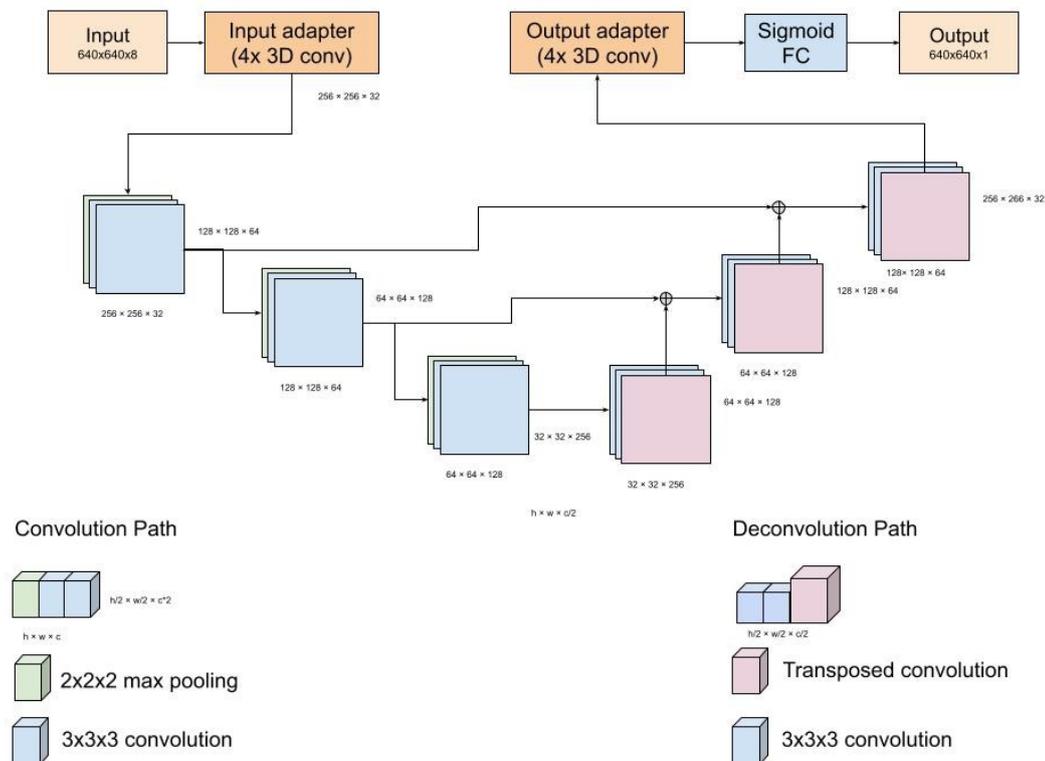


Figure 2: The diagram of the proposed approach

6.4. Segmentator Training

When creating a set of data for the proposed one, it is important to take into account the nature of the task and the specificity of the data.

To detect explosive objects, the images must have a high resolution, which makes the use of orthophotomosaic segmentation techniques impractical. Recognition must take place at the level of individual drone images, which reduces dimensionality, but increases the number of individual samples that need to be labeled. The received data contains a high level of noise, which may have a signature similar to the target objects. Also, the dataset itself is unbalanced and contains a large number of "empty" images with noise and a small number of images with explosive objects. Taking

this into account, semi-supervised learning based on spatial consistency with proxy labeling and modified loss functions and learning modes are used in training the segmenters.

This paper considers the proxy labeling method based on the principle of smoothness and clustering. The idea is that when applying the clustering algorithm, all the pixels belonging to the landmine class should belong to the same cluster, which allows the class label to be propagated within the same image. The disadvantage of this approach is that the propagation is sensitive to noise, and the transfer of labels between different images requires the application of complex cluster similarity metrics, which significantly reduces the accuracy of pseudolabels. Therefore, in this work, to compensate for these shortcomings, an additional assumption is introduced - temporal-spatial cluster consistency. The intuition is that by having a labeled image x_t and the unlabeled image x'_{t+1} , which have a certain intersection, both contain the object of interest. If the direction of optical flow is known, an approximate location of the object of interest in x'_{t+1} can be derived by offsetting the location of the object in x_t via the estimation of the camera's movement. Therefore, a distribution of the possible location of the object in the image with an estimation of the probability based on the distance is created. Afterwards, clustering of the image is performed and the intersection between the detected clusters and the object's location probability distribution is calculated, which is used to verify the accuracy of the pseudo labels. Visually, the idea is depicted in Fig. 3

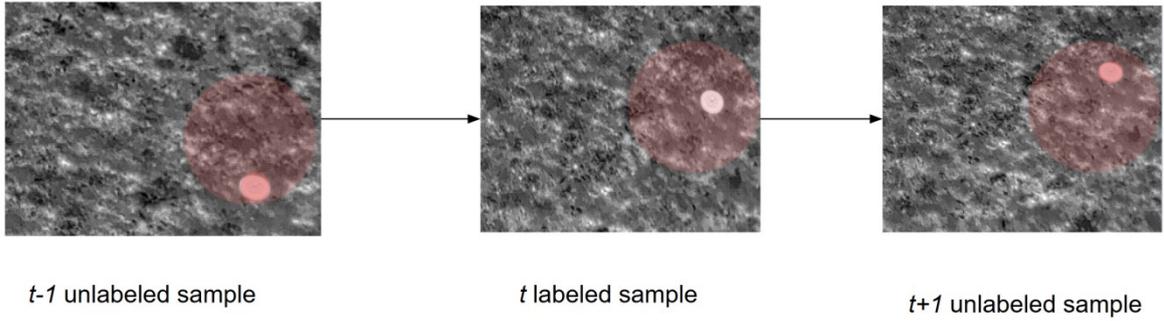


Figure 3: Semi-supervised learning with temporal and spatial consistency

Analytically, the procedure is moderately complex. The first step is the calculation of the probability map of object's location. This is achieved by applying a convolution operator with a Gaussian kernel is used, which is defined as:

$$K_{Gaussian}(x, y) = \frac{e^{-\frac{(x^2 + y^2)}{2\sigma^2}}}{2\pi\sigma^2}, \quad (3)$$

This kernel is applied to the labeled image x_t in order to generate a probability map p_t . The next step is to cluster the data of the unlabeled image x'_{t+1} . Any algorithm can be used for clustering, KNN is used in this work. The clustering algorithm is applied for each spectral channel, forming cluster masks $m_{t-1} \dots m_{t,c}$.

After calculating the cluster masks for each spectrum, the metric of the ratio of the noise to the positive signal of the label of each of the clusters in the image is calculated:

$$LS_{m_k}(x, y) = \frac{\sum_{x, y \in k} \sum_{w \in k - Gaussian(x, y)} W * X_t(x, y)}{|k|}, \quad (4)$$

where k – is a cluster of mt , w are pixel's weight in the Gaussian kernel, x_i is the labeled image. The value of SC_i is bounded by $[0, 1]$ and does not require additional normalization.

The noise ratio is calculated for each of the spectral channels, after which the spectral consistency metric is calculated, on the basis of which a conclusion is made regarding the assignment of the label to each of the pixels. Landmine features may be clearly present in only one (or several) spectral channels, as such logarithmic transformation is used to significantly increase the influence of high-impact channels on the pseudo-labeling results. Spectral consistency is calculated as:

$$SC_i(x, y) = \frac{\sum \log(1 + \alpha * LS_{m,k}(x, y))}{|c| * \log(1 + \alpha)}, \quad (5)$$

where c is the number of spectral channels, α is a scaling hyperparameter, recommended value is $[10, 1000]$, depending on the number of channels.

The last step of the algorithm is to cut off low-probability labels. For this, a standard cut-off process is used based on the ConfidenceThreshold parameter (CT). Pseudolabels with a value of spectral consistency less than CT are filtered out and not used in further training.

The pseudo-labeling is performed iteratively. For each labeled image, where a landmine segment is present, a breadth first search is used to check neighboring images in the order of optical flow and propagate landmine segmentation pseudo labels. The propagation is performed until the average value of spectral consistency SC_i is above the confidence threshold. This iterative process is repeated for every labeled image with landmine segment present.

An important component of system training is sensitivity adjustment during segmenter training. The following features must be taken into account during training:

- explosive ordnance are a minority class in the data set, which leads to class imbalance;
- one of the models should have high sensitivity and low specificity, the other one should have medium sensitivity and high specificity, while the models have a dozen architecture and learning method.

Taking into account these conditions, the work uses two approaches to training segmenters. In both cases, the same loss function is used – the binary weighted cross-entropy:

$$L_{WBCE} = -E[W_1 \cdot y_{pred}] \cdot \log(y_{pred}) + W_0 \cdot (1 - y_{true}) \cdot \log(1 - y_{pred}), \quad (6)$$

where w_1, w_0 are the weights of positive and negative class correspondingly, y_{pred} – is the predicted value, y_{true} – is the label value.

When training a sensitive segmenter, more aggressive parameters - $w_1 = 0.8, w_0 = 0.2$ - are used, which results in sensitivity loss, but in turn increases the recall. The model training is performed on an unbalanced dataset using batch-normalization.

When training a more specific classifier, less aggressive weights are used - $w_1 = 0.6, w_0 = 0.4$ - which increases the precision of this classifier. However, during training, resampling is used, which ensures the presence of explosive ordnance in 50% of the samples in the batch. This method of resampling balances the training and allows to achieve high accuracy and specificity for the validator sensor.

When training neural networks, the Adam optimizer is used, the learning rate parameter is set to 0.001 with a stepwise decay and a minimum value of 0.0001.

7. Results

The verification of the applied method was carried out by analyzing the effectiveness of the application of the intelligent system based on experimental data. The experiment was conducted using two agents: a drone with a multispectral camera and a cart with an infrared camera.

Data for the experiment were collected in the Khmelnytskyi region. TM-62 mine simulators were installed on the site with a total area of 0.5 square kilometers. When the mines were installed, two possible configurations were considered: installation directly on the surface and installation at a depth of up to 10 centimeters. After installation, an unmanned aerial vehicle with a multispectral camera was flown over, followed by a tour of places with known mine installations using a cart with an infrared camera. The collection was carried out in the summer, in warm (up to +23 degrees Celsius per day), sunny weather, which improves the quality of data collected using an infrared camera. The grass cover was present, partially disturbed at the place where the mines were placed.

Two types of cameras were used for data collection – the multispectral camera built into the Mavic 3M drone captures 5 multispectral channels and the visible spectrum. The flight was carried out at a height of 10 meters above the surface level. For the ground drone, a ZH20T thermal camera mounted on a cart was used. The height of the camera during filming was up to 20 centimeters. To capture footage, the camera must be directed perpendicular to the surface of the earth, which makes navigation difficult.

Figure 4: Landmine imagery collection area



Pseudolabeling significantly increases the number of labeled samples without using a large number of resources for labeling. As a result of marking, it was found that the marking efficiency strongly depends on the features of the sensor. Thus, multispectral data contains several channels, which makes spectral consistency an effective metric, while infrared data contains only 1 channel, which significantly reduces the effectiveness of the consistency metric and requires a higher threshold CT . The results of the pseudolabels used when initializing the segmenters are shown in Table 1.

Table 1

Pseudolabeling results

Dataset	Labeled	Unlabeled	Pseudolabeled
Multispectral	20	84	67 (79%)
Infrared	30	375	206 (55%)

After pseudolabeling, models were trained in cross-validation mode, with a 90-10% split into training and validation datasets. The confusion matrices of the segmentators are presented in Table 2, and the general learning metrics are presented in Table 3.

Table 2

Confusion Matrices (Full datasets)

Multispectral	Positive	Negative
True	10	69
False	8	0
Infrared		
True	26	206
False	3	1

Table 3

Training metrics(Validation Only)

Type	Recall	Precision	Rand	IoU
Multispectral	1.0	0.694	0.893	0.8385
Infrared	0.92	0.93	0.961	0.948

Overall, the approach works well, however semi-supervised learning could be tuned further. In table 2, it can be seen that the validator agent has 1 false negative, which is not acceptable for real-world condition. This false-negative was traced to a blurry image in the vicinity of a labeled landmine, which was not properly labeled in the frame. This issue was resolved by just removing the blurry image from a training set, however to adapt the method to the fluctuations in the input data either a regularization or data filtering techniques should be explored further.

Conclusions

The proposed method allows to train two classifiers using the unified framework. The semi-supervised methodology boosts sample size by ~3 times for datasets used in the study. Achieved results are in line with state-of-the-art methods and provides high recall and moderate precision at 0.694, which is further refined through the verification agent to achieve 0.93 precision.

The proposed method, however, is vulnerable to noise in the data which creates challenges during the semi-supervised learning. As such, future research will be focused on stabilizing the semi-supervised learning either by adding extra filters to the input data to detect, remove or pre-process low-quality samples or use regularization to ensure stable learning.

Future research will be focused on improving upon two key parts of the proposed method, namely semi-supervised learning for pseudolabeling and sensitivity training for different agents, Pseudolabeling part struggles with rapid shifts of the scene, especially if motion blur is introduced. Adding regularization parameter should stabilize the learning and improve generalization of the semi-supervised step. When it comes to the sensitivity training for different agents, current approach is able to utilize the same network architecture for both agents, with sensitivity tuning for each agent independently. This leads to longer training time, which could be reduced by adjusting training routine to re-use weights learned by one of the agents to train a second agent.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

- [1] Ismail, A., Elmogy, M., and ElBakry, H., "Landmines detection using low-cost multisensory mobile robot," *Journal of Convergence Information Technology*, 2015. 10(6): p. 51.
- [2] H. G. Najjaran, "Using genetic algorithms and neural networks for surface land mine detection," *IEEE Transactions on Signal Processing*, 47, pp.176-186, 1999.
- [3] Moamen Mohamed, et al. Autonomous Landmine Detection Robot Using SLAM navigation Algorithm. *Journal of International Society for Science and Engineering*, Vol. 4, No. 4, 81-91 (2022).
- [4] Anuar, A. Youssef, S. Said, I. Chuan, "The Development of an Autonomous Personal Mobile Robot System for Land Mines Detection on Uneven Terrain", *An Experience Conference on Intelligent Systems and Robotics*, u-tokyo.ac.jp, 2003.
- [5] P.Gonzalez de Santos,E. Garcia,J. Estremera and M. Armada," A. DYLEMA: Using walking robots for landmine detection and location "International Journal of Systems Science, Informa UK Limited, 36, PP.545-558 ,2005.
- [6] L.Robledo,M.Carrasco and D.Mery," A survey of land mine detection technology", *International Journal of Remote Sensing*,30,PP. 2399-2410,2009.
- [7] J.NICOUD and M.HABIB, 1995, "The Pemex-B autonomous demining robot: Perception and Navigation strategies", *Proceedings of the International Conference on Intelligent Robots and Systems*, 1, pp. 419–424, 1995.
- [8] P.Santana, J. Barata, H.Cruz, A.Mestre,J.Lisboa,andI.Flores,"A multirobot system for landmine detection", *10th IEEE Conference on Emerging Technologies and Factory Automation, ETFA*, Vol. 1, p. 8,2005.
- [9] Mats, H. Norwegian peoples IRAQ drone use and lessons learned. *Norwegian People's Aid Workshop on Lessons Learned from the Use of Unmanned Aerial Vehicles for the Identification and Assessment of Explosive Devices Threats*, Presentation, Podgorica, Montenegro, 16–17 October 2019.

- [10] Mats, H. Video acquired by RPAS. In Norwegian People's Aid Workshop on Lessons Learned from the Use of Unmanned Aerial Vehicles for the Identification and Assessment of Explosive Devices Threats, Podgorica, Montenegro, 16–17 October 2019.
- [11] Toolbox Implementation for Removal of Anti-Personnel Mines, Sub-Munitions and UXO—TIRAMISU, EU FP7 Project 2012–2015, Grant Agreement Number 284747. Available online: <http://www.fp7-tiramisu.eu/> (accessed on 9 January 2021).
- [12] Fardoulis, J. Drones in HMA lessons from the field 2019. In Proceedings of the 7th Mine Action Technology Workshop, GCIHD, Basel, Switzerland, 7–8 November 2019.
- [13] Nevard, M.; Mansel, R.; Torbet, N. Use of aerial imagery in urban survey & use of RPASs in mine Action—Lessons learned from six countries. In Proceedings of the 7th Mine Action Technology Workshop, GCIHD, Basel, Switzerland, 7–8 November 2019.
- [14] Lisica, D. Evaluation on use of UAVs in-country assessment of suspected hazardous areas in Bosnia and Herzegovina 2019. In Proceedings of the Norwegian People's Aid Workshop on Lessons Learned from the Use of Unmanned Aerial Vehicles for the Identification and Assessment of Explosive Devices Threats, Podgorica, Montenegro, 16–17 October 2019.
- [15] Sineglazov, V.M., Ischenko, V.P. Integrated navigation complex of UAV on basis of flight controller. 2015 IEEE 3rd International Conference Actual Problems of Unmanned Aerial Vehicles Developments, APUAVD 2015 - Proceedings, страницы 20–25, 7346547, 2015.
- [16] Sineglazov, V.M. Computer aided-design problems of unmanned aerial vehicles integrated navigation complexes. 2014 IEEE 3rd International Conference on Methods and Systems of Navigation and Motion Control, MSNMC 2014 - Proceedings, страницы 9–14, 6979716, 2014.
- [17] Kamarudin, K., Mamduh, S. M., Shakaff, A. Y. M., Saad, S. M., Zakaria, A., Abdullah, A. H., and Kamarudin, L. M. "Method to convert Kinect's 3D depth data to a 2D map for indoor SLAM." in 2013 IEEE 9th international colloquium on signal processing and its applications.
- [18] Kneip, L., Твче, F., Caprari, G., and Siegwart, R. "Characterization of the compact Hokuyo URG-04LX 2D laser range scanner." in 2009 IEEE International Conference on Robotics and Automation.
- [19] Gibson J. M., Lockwood J. R. Alternatives for Landmine Detection. – Santa Monica, CA: RAND Corporation, 2003. – XXX, 336 p.
- [20] Zgurovsky, M., Sineglazov, V., Chumachenko, E. (2021). Classification and Analysis of Multicriteria Optimization Methods. In: Artificial Intelligence Systems Based on Hybrid Neural Networks. Studies in Computational Intelligence, vol 904. Springer, Cham. https://doi.org/10.1007/978-3-030-48453-8_2
- [21] Zgurovsky, M., Sineglazov, V., Chumachenko, E. (2021). Formation of Hybrid Artificial Neural Networks Topologies. In: Artificial Intelligence Systems Based on Hybrid Neural Networks. Studies in Computational Intelligence, vol 904. Springer, Cham. https://doi.org/10.1007/978-3-030-48453-8_3

- [22] Li N., Wang Z. Cheikh F. A. Discriminating Spectral–Spatial Feature Extraction for Hyperspectral Image Classification: A Review // MDPI Sensors. – 2024. – 24, 2987. – <https://doi.org/10.3390/s24102987>. – 32 p.