Shallow Buried Improvised Explosive Device Detection Via Convolutional Neural Networks

Simon Colreavy-Donnelly [a], Fabio Caraffini [a,*], Stefan Kuhn [b], Mario Gongora [a],
Johana Florez-Lozano [c], Carlos Parra [c]

[a] *Institute of Artificial Intelligence, School of Computer Science and Informatics, De Montfort University, UK*
E-mails: simon.colreavy-donnelly@dmu.ac.uk, fabio.caraffini@dmu.ac.uk, mgongora@dmu.ac.uk
[b] *Cyber Technology Institute, School of Computer Science and Informatics, De Montfort University, UK*
E-mail: stefan.kuhn@dmu.ac.uk
[c] *Pontificia Universidad Javeriana, Carrera 7 No. 40 - 62, Bogota, Colombia*
E-mails: johana.florez@javeriana.edu.co, carlos.parra@javeriana.edu.co

Abstract. The issue of detecting improvised explosive devices, henceforth IEDs, in rural or built-up urban environments is a persistent and serious concern for governments in the developing world. In many cases, such devices are plastic, or varied metallic objects containing rudimentary explosives, which are not visible to the naked eye and are difficult to detect autonomously. The most effective strategy for detecting land mines also happens to be the most dangerous. This paper intends to leverage the use of a Convolutional Neural Network (CNN) to aid in the discovery of such IEDs. As part of a related project, an autonomous sensor array was used to detect the devices in terrains too hazardous for a human to survey. This paper presents a CNN and its training methodology, suitable to make use of the sensor system. This convolutional neural network can accurately distinguish between a potential IED and surrounding undergrowth and natural features of the environment in real-time. The training methodology enabled the CNN to successfully recognise the IEDs with an accuracy of 98.7%, in well-lit conditions. The results are evaluated against other convolutional neural systems as well as against a deterministic algorithm, showing that the proposed CNN outperforms its competitors including the deterministic method.

Keywords: land mine detection, convolutional neural network, land sensing, image processing, improvised explosive device

1. Introduction

Detecting improvised explosive devices in urban and rural environments is a new and challenging field of study that has great potential, in terms of humanitarian developments in Robotics [1–3], machine learning [4, 5] and Chemical Engineering [6, 7].

An Improvised Explosive Device (IED) can vary from simple plastic containers to metallic objects containing explosives. This makes them difficult to recognise and distinguish from surrounding terrain, particularly if they are subterranean, partially-buried or well disguised in the surrounding scrub, rubble or sand. Examples of improvised mines built with everyday items, e.g. soft drink bottles, PVC pipes, syringes, are shown in figure 2. This paper presents the use of a Convolutional Neural Network (CNN) [8], to aid the discovery of such IEDs, by examining features like heat, shininess and shape in image analysis. A CNN was trained to detect these devices. The network is able to distinguish the devices, by looking at heat markers and specularity, specifically; often IED's have telltale heat signatures, which give them away.

The resulting CNN is intended to be deployed in a detection platform and could prospectively be used for mine clearance, or removal of hazardous materials. This paper shows how this CNN can accurately distinguish between a potential IED and surrounding undergrowth and natural features of the environment in real-time.

---

*Corresponding author. E-mail: fabio.caraffini@dmu.ac.uk.

The main objectives and challenges of this piece of research are outlined in section 1.1. Following, a thorough literature review is provided in section 2 to give a wider overview of machine learning approaches for detection purposes. Material and methods used to pursue our research goal are described in section 3. For the sake of reproducibility, this section also contains technical details and a description of the procedures designed to obtain and evaluate results. The obtained results are reported and commented in section 4 and conclusions on this study are drawn in section 5.

1.1. Objectives and motivations

The focus of this article is on detecting improvised mines with the help of a Deep Learning approach using Convolutions Neural Networks. The latter found their use in a variety of application domains, mainly image recognition [9] and several other fields [10–13], but to the best of our knowledge, they are yet to be exploited for the detection of IED's. This extends the successful CNN application for various mine clearance scenarios shown in section 2.

However, due to the diverse nature of improvised land mines, systemic feature recognition and determination are quite difficult and significantly more challenging than detecting military mines. A prospective Deep Learning solution for this problem is also quite pertinent, given the suitability of Deep Learning models and Neural Systems for feature recognition and determination, considering specifically features that contribute to give away the position of land mines (heat, specularity, shininess and so forth). This has resulted in the choice of a Convolutional Neural Network to provide a solution to this problem. It must be added that this choice allows for the detection of patterns in the ground which even a GPR may miss given that, unlike military mines, most improvised explosives do not have an electromagnetic signal. Furthermore, it can be said that image processing methods are more affordable, practical and much faster to deploy than classic methods based purely on GPR data.

2. Literature Review

The problem of detecting mines in an image can be considered an image segmentation problem. Simple methods used for this are thresholding, clustering, or histogram-based methods. These segment the image into regions of similar colour, hue, or saturation. Re-

lated are methods based on edge detection, which assume that edges separating segments show particular characteristics. Over time, many different variants and combinations of these methods have been suggested, as e.g. those in [14] and [15]. For an overview and literature, see [16]. A next step is a semantic segmentation, where areas of the image are not only identified, but also labelled as containing a certain class of object. Instance segmentation, in addition, separates the potentially multiple instances of an object. Traditional approaches (i.e. those not using machine learning methods) mostly use a sliding window algorithm, using different combinations of features to look at. This has been combined with various other techniques like dimensionality reduction or manually added features. See [17] for a survey. We will call those methods deterministic in this paper.

Various machine learning techniques have been applied to the problem of image segmentation and semantic segmentation. They include Support Vector Machines (SVMs) [18] and Random Forest [19]. Artificial Neural Networks (ANN) are widely used for image segmentation, for example, Kohonen self-organising maps [20] and other "shallow" neural networks.

Deep Learning neural systems, being the latest incarnation of machine learning and currently considered state of the art in the field, have been investigated and applied in different domains such as civil engineering [21–23], data clustering [24, 25] and many other fields [10, 26–33]. These research activities led to significant improvements also in image segmentation [15, 34]. An interesting CNN approach is taken in [35], where material properties are detected. However, the most prominent implementations are those which have won the ImageNet competitions over the past years: AlexNet [36] in 2012, VGG-16 [37] in 2013, GoogLeNet [38] in 2014 and ResNet [39] in 2016. Typical application areas include autonomous cars [40–42] and medical image analysis [34, 43, 44]. The latter was the aim of the Unet fully convolution network architecture [45]. Other medical applications where Deep Learning neural systems have been found to play a major role are signal analysis [46, 47], e.g. for interpreting encephalograms, and providing diagnosis [48–50].

Specifically for the problem of land mine detection both deterministic and machine learning methods have been employed. In [4] a method was used that detects mines using a Ground Penetration Radar (GPR). It should be noted that this method is deterministic,

rather than entirely Deep Learning based. Soil analysis and GPR has been a specific approach to land mine detection, e. g. [51]. The studies in [52] represent the most comprehensive account of land-mine detection topics; the book collects papers that were presented at the 2008 NATO Advanced Study Institute on Unexploded Ordnance Detection and Mitigation. Another review of land mine detection methods can be found in [53]. In [54], the authors proposed a cellular neural network (which must not be confused with a CNN) for the 3D thermal modelling of soil. [55] makes use of a CNN for spotting patterns in thermal images of buried items. Other studies using neural networks to analyse GPR data can be found in [56], which proposes an ANN approach and a fuzzy approach, and [57], where a Fast R-CNN is employed. Another CNN-based approach using GPR data can be found in [58]. These neural network-based approaches were particularly effective at recognising mines, but the authors were more focused on recognising military-standard mines, rather than the improvised type looked at in this paper. Finally, in [59], a novel CNN method for detecting mine-like objects using sonar was presented. It should be noted that this method is deployed in the sea, rather than on land. All the comparative methodologies using subtly different techniques from what is proposed in this paper and this shall be discussed further in section 4.

3. Material and Methods

3.1. Image acquisition

The multi-device platform in [60] was used to collect images and generate the inputs for the proposed CNN. As displayed in figure 1, it is equipped with cameras to scan the terrain. A "Nikon D5300" camera with a 55mm lens is used take RGB photographs. Conversely, two "Nikon D5200" with a 55mm lens are used together with an infrared filter and an ultraviolet filter respectively, plus a ring illuminator, to obtain photographs in those frequency bands. Each camera runs the AI algorithms available from [61] to self-adjusts its angular position, so to maximise the quality of the information on each image acquired, which significantly overlaps with the next sample. This allows for an efficient image registration process and accurate reconstruction of the test terrain in a single high-resolution image, which still contains the detailed information of every single original image.
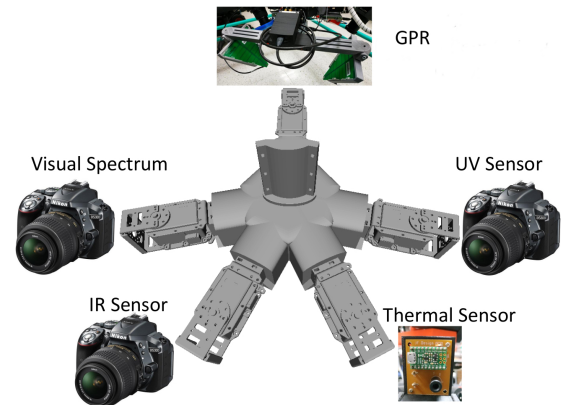

Fig. 1. Sensors of the improvised land mine detection platform in [60].

Three mock improvised mines from those shown in figure 2 were buried in a lot of size 670×1100 mm to safely perform the IED detection task.

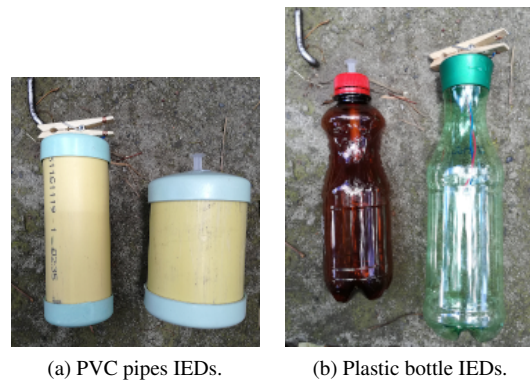(a) PVC pipes IEDs.　　(b) Plastic bottle IEDs.

Fig. 2. Mock improvised land mines used in the project [62]. More images (source high resolution files) are available in [60].

The masks are shown in figure 5 indicates their location, which can be given in Cartesian coordinate as $(550, 250)$, $(350, 600)$ and $(500, 850)$, where the origin of the coordinate systems is the bottom left corner of test terrain, the *x* and *y* axis are oriented towards the right-hand side and up respectively, and distances are measured in mm. To reproduce real conditions, IEDs were buried months before the acquisition phase to let the vegetation grow back, and the test terrain was protected from city debris that is not usually present in rural lands, where war and conflict take place.

A significant amount of images are acquired with each camera while scanning the test terrain using the Cartesian robot in figure 3. Hence, the latter is config-

Fig. 3. Improvised land mine detection platform in action. For testing purposes, the test terrain is scanned via a Cartesian robot.

ured to move the cameras with a fixed step of 50 mm to have sufficient images to be "stitched" together by aligning the coordinates of each sample, thus generating the inputs for the CNN, shown in figure 5 and 6. A complete collection of the generated input images used in this study can be found in the repository [63].

Furthermore, it is worth pointing out that data collected with the remaining sensors, available in [60], have also been processed in previous attempts but led to poor detection performances when considered individually. Therefore we focused our attention on processing images and decided to use CNNs given their success in manipulating such inputs.

These images, generated as part of our ongoing research, are published in a repository [61]. Later in this paper, a separate dataset is used from another project [64], as explained in section 3.4.

3.2. The neural detection system

The VGG-16 network architecture was used as a basis for the mine-detection CNN [65]. It was altered to take direct inputs from the infrared or ultraviolet filtered images in addition to the visual channel, incorporated in the multi-device platform. This (infrared/ultraviolet) input is rendered into a 2-D probability/greyscale distribution in an additional layer. The comparison of these channels with the original RGB image and a black and white 1-D label mask form the input layers and validation processes of the network: effectively comparing noisy/labelled inputs provides greater accuracy than labelled data alone. This is especially the case with the improvised IED examples, where there are no specific tell-tale electromagnetic signals to train a CNN with. Instead, only noisy image-

based data can be used to train the CNN. These layers of the CNN are shown in figure 4. Rectified Linear Units (RELU) described in [66], i.e. the de-facto standard threshold algorithm in image classification and segmentation in Deep Learning [36, 67, 68], is used as activation function for the CNN. A grid search is used to determine the activation function (RELU, Softmax, Sigmoid), the number of layers and number of neurons per-layer. This neural system was implemented in Python, and common machine learning modules as [69–71] were used. Its structure presents a 32×32 convolutional filter, an RGB image input and one for the infrared/ultraviolet filters, with validation provided by the black and white label-mask. Due to the impossibility of paralleling the execution of the network, as the system is designed to be deployed in a mobile autonomous platform, training was performed on a CPU. This must be taken into account when looking at the results of the testing, reported in section 4, in terms of both training and testing elapsed time (displayed in table 1). The reason for the implementation of the infrared and ultraviolet filters is two-fold: firstly, the infrared channel may be able to detect the heat signature of an IED buried (or partially-buried) under the ground and secondly, the ultraviolet channel may detect mines in low light conditions. Another important point is that an IED made of plastic or metal will have higher specularity, so the ultraviolet channel may pick this up also when employed in combination with infrared.

3.3. Training

Training inputs are the RGB image input and two other inputs, one for the infrared/ultraviolet filter and validation provided by the black and white label-mask. The RGB/UV/IR training images are composed of two-dozen individual images taken with a digital camera, which are joined together in the final images shown in figure 5 and 6, which form our final data set. This is a purely image-based approach, which can be executed on the platform, in the field, since it is lightweight enough to be executed on an embedded system. These images are then passed through the convolution filters of the network, testing the aggregate infrared and ultraviolet probability map channels with the black and white validation labels. Two sets of data were used for completeness. One set featured *light* imagery and the other an *overcast* landscape, to provide variety of training. The training data employed in the implementation of the CNN necessarily involved many false-positive examples, or patches that had several dif-
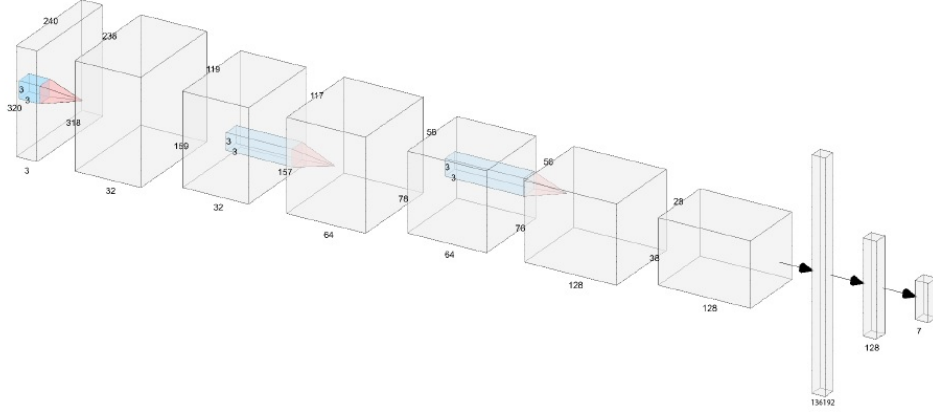
Fig. 4. The architecture of the employed Convolutional Neural Network. RELU, in this case, is the chosen activation function. The CNN consists of 32×32 convolution filters and it is structured as follows: input layers, a visual (RGB image), the infrared/ultraviolet filter of a digital camera, rendered into a 2-D grayscale probability distribution and compared to a 1-D black and white label, fed to a fully-connected layer of the CNN.

ferent examples of terrain feature (grass etc.), but no actual improvised mine. The patches which had improvised mines were marked out and the actual IED predictions were labelled with white pixels, as shown in the ground-truth/output comparison images in figure 5 and 6. The optimal number of training epochs to maximise performance by avoiding overfitting, as well as underfitting, was found experimentally to be 600.

3.4. Validation

To strengthen the validation phase, the additional dataset from [64] was employed. The latter was carefully selected considering that 1) CNN systems should not be assessed on the same dataset that used for their training; 2) the adverse environmental conditions in the training dataset [63] make it very difficult to find several other datasets for performing a "like-for-like" comparison in the literature; 3) the proposed system itself will more than likely always be deployed on dissimilar datasets. Furthermore, we use a deterministic sliding-window algorithm for validation. By applying this to the same images as the CNN we can establish the actual advantages of the machine learning method.

3.5. Deterministic image analysis

The deterministic sliding-window algorithm detects specular, shiny or reflective objects by looking at changes in regional variation between hue and intensity; the original area where this algorithm was employed was in video processing [72]. For IEDs, areas

where hue does not change significantly, but the intensity does, are good candidates for a non-reflective (Non-IED) region; conversely, areas, where the opposite is true, are good candidates for a shiny, reflective (IED) surface, potentially indicating the presence of an IED. The sliding window/image is calculated through equation 1 as follows:

$$\text{Non-IED} = \begin{cases} 1, & \text{if} \quad \overline{\text{H}}_w > \overline{\text{H}}_i \vee \overline{\text{V}}_w > \overline{\text{V}}_i \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where *Non-IED* is a non-reflective candidate pixel, and its value is black (i.e. 1), if the mean hue of the window $\overline{\text{H}}_w$ is greater than the mean hue of the image $\overline{\text{H}}_i$ *and* the mean intensity of the window $\overline{\text{V}}_w$ is greater than the mean intensity of the image $\overline{\text{V}}_i$, otherwise it is white (i.e. 0) and may be a candidate for an IED. The sliding window is set to dimensions of 11×11 pixels, with a centre pixel-value, from where the window moves. The mean i, j^{th} pixel-values of the window are calculated with equations (2), for the hue and intensity of their respective positions:

$$\overline{\text{H}}_w = \frac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} H_w ij}{n^2}, \quad \overline{\text{V}}_w = \frac{\sum\limits_{i=1}^{n} \sum\limits_{j=1}^{n} V_w ij}{n^2}. \tag{2}$$

The window values are then compared, at these positions, with global hue and intensity features, as noted in equation (2).

There are two important reasons for comparing the CNN against a deterministic algorithm. Firstly, due to the hazardous nature of training a robot in real-time in a minefield, the extent of the data available is limited. Much of our training data is negative data, rather than actual positive examples of the explosive improvised mines. For a Deep Learning approach, this can be problematic. However, a deterministic method not based on a learning process is helpful to benchmark against. This is because a deterministic algorithm will either entirely fail or succeed in categorising pixel regions of an image regardless of training, whereas a Deep Learning solution involves an iterative process. Usually, a deterministic methodology will outperform a Deep Learning method at the initial stages of training or calibration. Secondly, there is a limited number of Deep Learning based solutions to the mine-detection problem, so the deterministic results allow for a direct comparison with our methodology and data set. As mentioned in section 2 this is work done using machine learning methods [4, 54, 59], but they generally focus on different types of data, such as underwater mines or are from older studies.

3.6. Comparative Analysis

To enable a fair comparison, we first compare the results obtained with our proposed CNN to the results of the deterministic method over the data collected with our system, as explained in section 3.1, and then we further compare the results obtained with our proposed CNN method to those from established similar methods using the dataset in [64]. In the last comparison with the state-of-the-art, accuracy, Intersection Over Union (IOU) [73], and Sørensen-Dice Coefficient (SDC) [74] are employed as evaluation metrics.

For calculating the accuracy of the CNN the semantic segmentation pixel-counting method from [75] was employed, whereby the number of correctly guessed black-pixels was averaged with the number of the correctly guessed white-pixels, with black pixels representing Non-IED pixels and white pixels representing IED pixels. Here, we count the number of pixels of those specific colours. For this reason, this metric is here referred to as Black/White (B/W) accuracy. Its definition is displayed in equation 3:

$$\text{B/W} = \frac{\text{IED}_{\text{accuracy}} + \text{Non-IED}_{\text{accuracy}}}{2} \tag{3}$$

where

$$\text{Non-IED}_{\text{accuracy}} = \frac{\text{Black Pixels}}{\text{Non-IED Pixels}}$$

$$\tag{4}$$

$$\text{IED}_{\text{accuracy}} = \frac{\text{White Pixels}}{\text{IED Pixels}}$$

are the accuracy for the IED and Non-IED regions respectively.

The IOU metric, aka Jaccard method, is a methodology whereby similarity and diversity of samples can be ascertained; it is used generally in image processing as a method to measure how far pixels match between two images. Its specific application is in image segmentation and classification. Hence, this is a suitable evaluation metric for the problem addressed in this study. The formula for calculating this metric is represented below:

$$J(A, B) = \begin{cases} 1 & \text{if } A = B = \emptyset \\ \frac{|A \cap B|}{|A \cup B|} & \text{otherwise} \end{cases} \tag{5}$$

with

$$\frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$

where A and B are two intersecting sets. In our case, the intersecting sets are the pixels of the ground truth compared with the pixels of the prediction, comparing the black/white pixel regions.

Finally, the Sørensen-Dice Coefficient differs from the previous metric in that IOU only counts true positives once in both the numerator and denominator. In this way, Sørensen-Dice coefficient gives a more general similarity measure. Its mathematical representation is shown below:

$$\text{DSC} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \tag{6}$$

where TP denotes true positive, FP denotes false positive and FN denotes false negative.

Two further important metrics which use these values and are themselves used in the calculation of Sørensen-Dice Coefficient, are Precision and Recall.
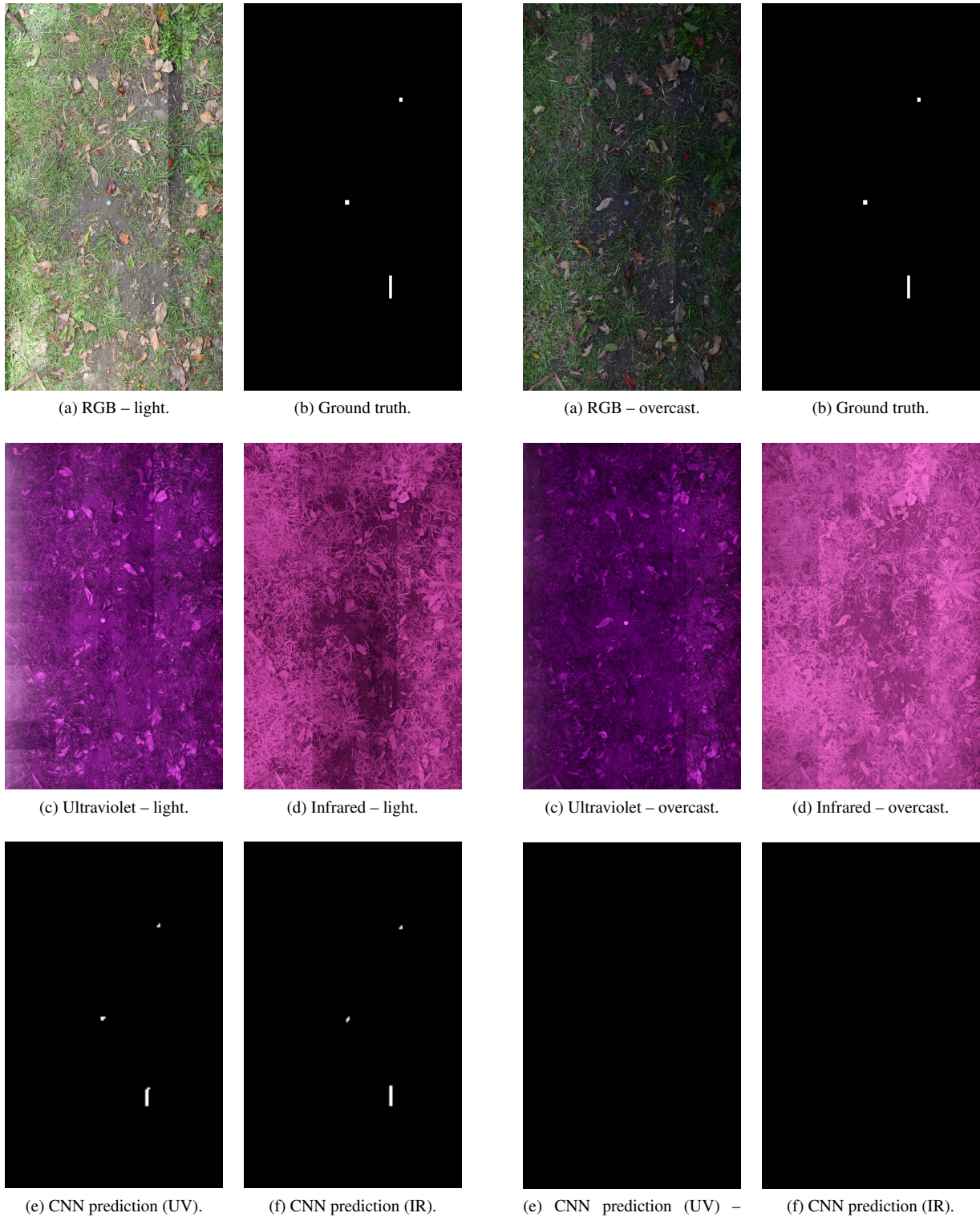
(a) RGB – light.

(b) Ground truth.

(c) Ultraviolet – light.

(d) Infrared – light.

(e) CNN prediction (UV).

(f) CNN prediction (IR).

Fig. 5. Predicted positions for the "light" scenario (a). The ground truth is reported in (b) to compare with the CNN binarised outputs obtained with the ultraviolet (c) and infrared (b) inputs, which are displayed in (e) and (f) respectively. UV and IR are short for ultraviolet and infrared.

(a) RGB – overcast.

(b) Ground truth.

(c) Ultraviolet – overcast.

(d) Infrared – overcast.

(e) CNN prediction (UV) – overcast.

(f) CNN prediction (IR).

Fig. 6. Predicted positions for the "overcast" scenario (a). The ground truth is reported in (b) to compare with the CNN binarised outputs obtained with the ultraviolet (c) and infrared (b) inputs, which are displayed in (e) and (f) respectively. UV and IR are short for ultraviolet and infrared.
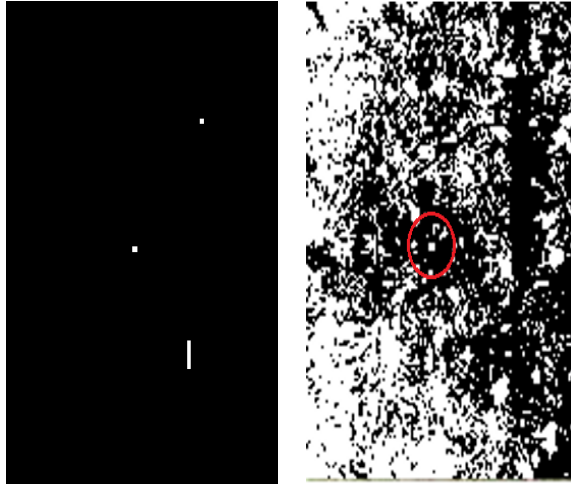
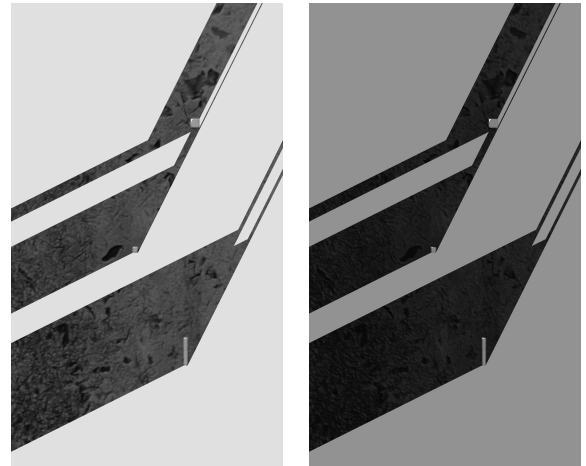(a) Ground truth.

(b) Deterministic.
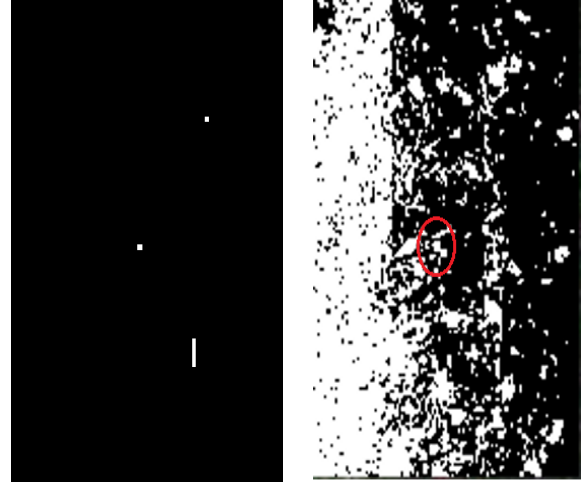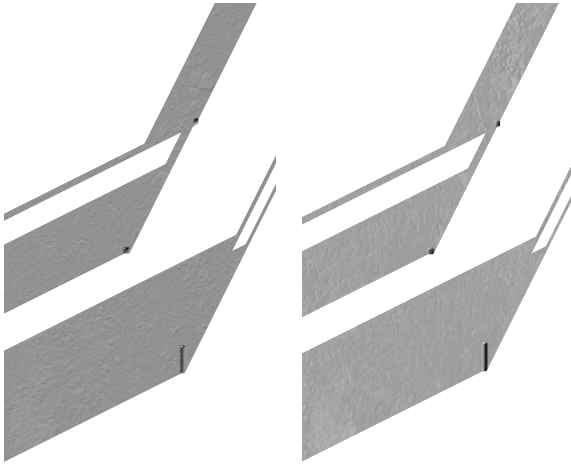
(c) CNN – ultraviolet.

(d) CNN – infrared.

Fig. 7. Comparison between the outputs obtained with the deterministic method (b) and the CNN approach, with ultraviolet (c) and infrared (d) outputs, for the "light" scenarios shown in figure 5a.

These are defined as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}; \qquad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}; \qquad (8)$$

from which it can be seen that Precision is a classification of the proportion of correctly guessed positive identification instances while Recall is a classification of the total proportion of actual positives, which was

(a) Ground truth.

(b) Deterministic.

(c) CNN – ultraviolet.

(d) CNN – infrared.

Fig. 8. Comparison between the outputs obtained with the deterministic method (b) and the CNN approach, with ultraviolet (c) and infrared (d) outputs, for the "overcast" scenarios shown in figure 6a

identified correctly. In this light, the DSC coeffieint can be expressed as the sum of the Precision and Reallmetrics.

4. Results

Graphical results are arranged as follows:

– figures 5 and 6 show the results of the CNN are after binarisation and post-processing;
– figures 7 and 8 show the actual outputs of both the CNN and deterministic algorithm respectively to facilitate their comparison.

The proposed CNN generated these results in 11.10 seconds (CPU).

The images show the entire plot (terrain) and are arranged into samples under two different environmental conditions, i.e. the light and overcast scenarios, shown in separate figures. These are placed side by side to facilitate their comparison. Hence, both figures 5 and 6 as well as figures 7 and 8 share the same structure, and show the same IEDs.

In the first case, the stitched RGB image of the test terrain is shown next to the ground truth in the top line to describe the experimental-setup. The corresponding ultraviolet and infrared filter images are displayed below and their predicted outputs appear in the bottom line of the figure. It must be clarified that the CNN produces as an output a greyscale probability map/distribution obtained from the input images shown in figure 5 and 6. This is the result of the network comparing and convolving the RGB input, the 2-D grayscale probability map (of the ultraviolet/infrared channel) and the black and white label mask.

Therefore, the predicted results from the CNN were thresholded using a means-average method: if the grey pixel-values fall below a prediction value of 28% they are black, as can be seen in figure (6e) and (6f), otherwise they are white, as shown in figure (5e) and (5f). The average of 28% was arrived at after the network was trained and tested several times on different data sets, using different metrics. All the results fell within a one or two % margin of this (28%) value. Most importantly, all results provided in this study used this same thresholding value for accuracy evaluation and comparison. We provisionally advise on using this value while using this neural network in future works with similar data sets. Further evaluation of the threshold, depending on conditions like daylight, is needed.

It should also be stated that the authors tried using the above thresholding value, i.e. 28%, to check if it could be applied to the raw greyscale images of the UV and IR channels, without having been processed by the CNN to detect the mines. The results were predictably inaccurate, achieving an accuracy significantly less than the aforementioned deterministic methodology, which was itself vastly inferior to the CNN.

In the second case, the ground truth is placed in the top left corner of figure 7, for the light scenario, and figure 8, for the overcast scenario. Next to it, the output of the employed deterministic algorithm is displayed, followed by the unprocessed (greyscale) outputs from the CNN for the ultraviolet and infrared inputs.

Methods	B/W (%)
IED CNN	98.7
CNN [59]	**99.0**
CNN [54]	95.0
CNN [58]	95.0
k-NN/Fuzzy [56]	98.0
GPR [4]	85.0
CNN-Fusion (AP) [64]	82.0
CNN-Fusion (AT) [64]	97.0

Table 1

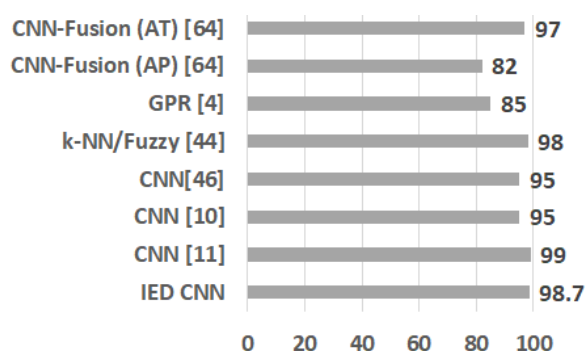Accuracy of the eight considered methods.

Fig. 9. Accuracy barchart of the eight considered methods.

Using the Accuracy B/W measure described in section 3.6 the CNN, used on the IED images, had an accuracy of 98.7%, which currently represents a benchmark in this area. The results, compared to other methods, are reproduced in table 1 and figure 9 for an overview.

As can be seen from the output, the CNN was clearly able to identify the IED's from surrounding regions, while the deterministic algorithm had much less success in this regard. It should be noted that these results are taken from the light scenario (figure 5), the results from the overcast scenario were much less conclusive, although as can be seen in the output in figure 8 the CNN could "see" the IED's, but had more difficulty distinguishing them from the surrounding terrain.

4.1. IED Dataset Comparative Analysis

The methods employed for comparison [4, 54, 59] use subtly different evaluation methodology to assess their accuracy. In [59] the accuracy was higher overall, at 99%, but this was for detecting mines in the sea, rather than on land. The other Neural Network method in [54] had an accuracy of 95%, as did in the one in [58]. The most accurate machine learning method for

land-based land mine detection was the "k-NN Fuzzy" hybrid method in [56], at 98.2%. However, this result was obtained from "active" military-grade land mines, rather than the more improvised materials examined in this paper (military-grade mines have known and consistent structure and materials so that sensors can be specialised and tuned to detect these). Finally, the method in [4] was a deterministic method, which is difficult to compare to the method in this paper, as the authors provide an accuracy ranging between 70% and 100% depending on the "noise" parameter in the GPR cross-section. Given that the data set used in this paper also contained significant amounts of noise (in the roughness of the terrain used for evaluation), the results were determined to be roughly an average of the "noisy" and "non-noisy" result, at 85%, though this value is of course not thorough.

4.2. Further Comparative Analysis

To complete our comparative analysis the performances of the proposed detection method are evaluated over different data set from similar studies, i.e. where CNN is used instead of a deterministic method. Since only a little is done for the IED problem, we selected two data sets, namely the indoor Anti-Personnel (AP) and indoor Anti-Tank (AT) mines data set respectively, from the study in [64]. These data, compared to our images, seem to be less challenging. First, the considered mines are usually military and not improvised. Secondly, a portion of the employed mines is often visible as the detonator cannot be completely hidden in the undergrowth. This is evident by inspecting figure 10a, which displays the type of mines used in [64]. However, the two data sets from [64] looked appropriate for our purpose since the methodology used for detecting AP and AT mines shares similarities to the one proposed in this article. Indeed, as graphically shown in figure 10, also in [64] labels of the mine position were fed to a CNN, along with multi-spectral data. We used the data from this study as the basis to do a comparative analysis. We processed the data using our CNN and then calculated the accuracy in terms of Accuracy B/W, IOU and SDC performance indices. For completeness, we have also used provided the Precision (Pr) and Recall (Re) metrics, used to calculate the SDC value. The obtained numerical values are shown in table 2 and in figure 11. Note that the numbers are shown correct to one decimal places and the values in Precision and Recall were actually slightly less than 100%. Its important to note that Precision and Recall

(a) Mine Position and Layout.

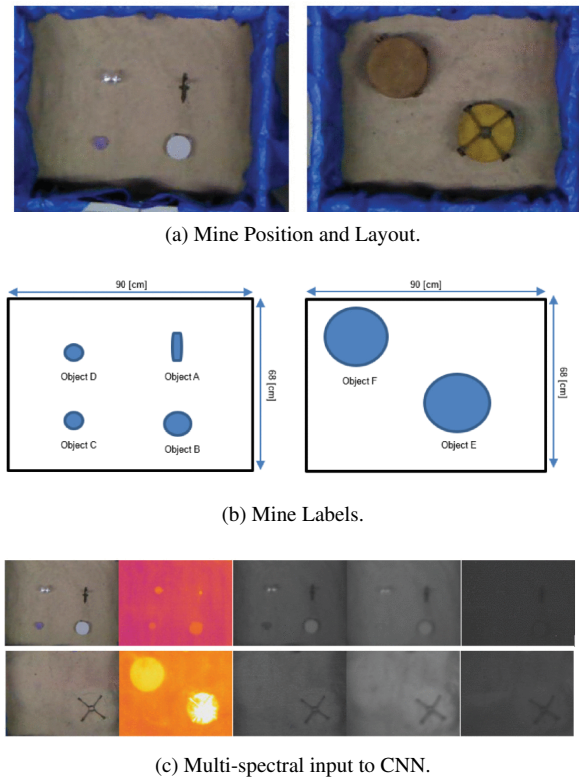(b) Mine Labels.

(c) Multi-spectral input to CNN.

Fig. 10. Overview of the methodology used in [64].

metrics are not particularly useful when considered on their own merit and are only useful when used in conjunction with SDC.

Data set	B/W	IOU	SDC	Pr	Re
IED [63]	98.7	77.7	89.0	100.0	100.0
AP [64]	98.7	96.0	98.0	93.0	100.0
AT [64]	**99.7**	**98.9**	**99.0**	**100.0**	**100.0**
Standard Deviation	0.60	14.8	4.49	3.29	0.0

Table 2

Performance of the proposed CNN over multiple data sets. IED is the proposed data set for improvised land mines available in the repository at [63]. AP and AT refer to the two data sets for military land mine detection from [64]. The highest accuracy value is marked in boldface and all values are given in per cent (%). The last line shows the standard deviation for each metric over the three datasets.

As can be seen from the above results, the proposed CNN performed significantly well when tested on the AP and AT data sets in all metrics of evaluation (with a better performance obtained over the AT data set). This was expected as, unlike the IED data set generated for this study [63], the indoor data sets in [64] are not noisy and the AP/AT mines are visible. This is high-
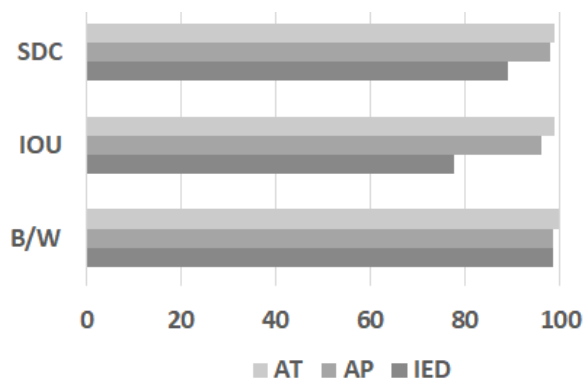
Fig. 11. Performance barchart of the proposed CNN on different data.

lighted by the fact that IOU and SDC are nearly as high as the B/W metric in these areas. As noted, the prevalence of noise in the IED data set may have been a factor in this. Finally, it is worth pointing out that the accuracy values obtained in the original study in [64] are 82% (AP data set) and 87% (AT data set), as reported in table 1, which means that our CNN implementation is competitive and outperforms other approaches also in the AP/AT application domain. Graphically, the obtained outputs for AP and AT cases are depicted in figure 12 and 13 respectively.

It should be further emphasised that the methodology used with the above data-set is to employ our own CNN, trained with our data-set, on the other data-set, using additional IOU and Sorenson metrics: we are not referencing their results, we are using their data-set to test our system. In contrast, the numbers in table 1 come from different studies, performed with different methods on different data. We also observed a significantly low variance, across the IED, AP and AT portions of the Silva et al. [64] dataset, making the total std deviation quite low, though difficult to determine, given the wide-ranging nature of the datasets.

5. Conclusions

In conclusion, this paper has explored several novel techniques in detecting improvised explosive devices, while putting forward its own new, enhanced IED CNN method. This work is of great significance because of its potential for developments in machine learning and for its humanitarian scope . This paper proposes the use of a convolutional neural network, to aid in discovery of IED's, by examining and extrapolating features like heat, shininess and shape from the

(a) RGB image.

(b) Ground truth.

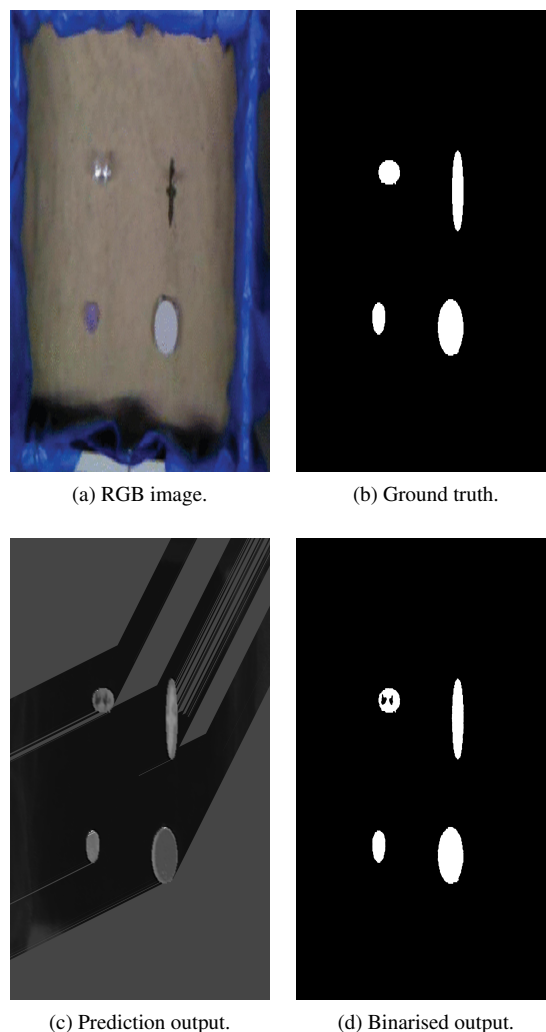(c) Prediction output.

(d) Binarised output.

Fig. 12. Graphical experimental results obtained with the proposed CNN over the AP data set from [64]. The first row displays the original RGB image and the corresponding ground truth in (a) and (b) respectively. The second row displays the predicted and the binarised outputs in (c) and (d) respectively.

infrared and ultraviolet channels of a digital image. To complement an autonomous robot which surveys dangerous terrains, the described CNN was successfully trained to detect the IEDs. The system can accurately distinguish the devices, by looking at heat markers and specularity.

The resulting CNN can be deployed in a robot and can be used for mine clearance, or removal of dangerous objects. The results show that this CNN can quickly and accurately distinguish between a potential IED and the surrounding terrain. The system has an accuracy of 98.7%, in non-overcast conditions, which is

(a) RGB image.

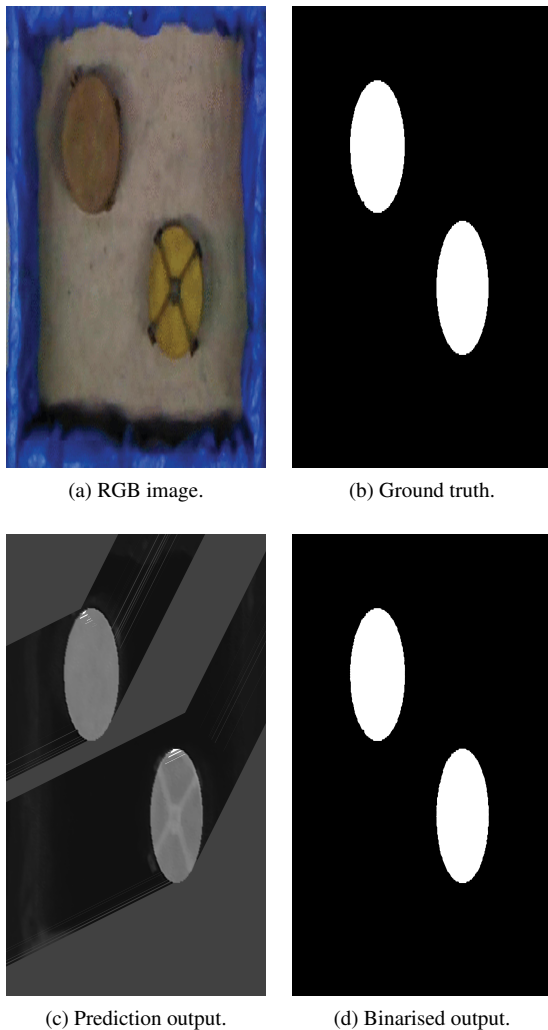(b) Ground truth.

(c) Prediction output.

(d) Binarised output.

Fig. 13. Graphical experimental results obtained with the proposed CNN over the AT data set from [64]. The first row displays the original RGB image and the corresponding ground truth in (a) and (b) respectively. The second row displays the predicted and the binarised outputs in (c) and (d) respectively.

currently a bench-mark for land-based mine detection. Significantly, the CNN was superior to the deterministic approaches examined, while comparing favourably with competing machine learning methods. It has to be noted that it was difficult to find methods with which to compare, as there is a limited number of Deep Learning based solutions to the real-time mine-detection problem [4, 52, 54, 59], but this is understandable, given the appreciably dangerous nature of land-mine detection and Deep Learning training in the field. The system is also the most computationally efficient of the methods examined, which is important when dealing

with live land mines in the field. As a means to extend and provide more accurate metrics for comparison and analysis, we also employed the Intersection Over Union and Sørensen-Dice Coefficient methodologies and a different data set, i.e. the one in employed in [64], to perform further comparisons and validate the potential of the proposed detection method. Results show that our CNN compared favourably with this data set. However, it is important to remark that [64] focuses on detecting military standard mines rather than IED's and this is an important difference, when taking into account accuracy and performance metrics.

Given that the proposed method appears to be promising, we intend to further investigate its use in the proposed platform. We will collect more images in different configurations of terrains and IED types to have a larger training data set. In particular, since this study is limited to the detection of shallow buried improvised mines (i.e. the detonator must be very close to the ground surface), we intend to expand to detect also improvised mines that can be buried deeper in the ground, as e.g. those activated remotely. In these cases, the denotation can occur when a remote detonation system, placed on the ground inside e.g. a piece of wood, is triggered by a pressure activated switch.

Furthermore, it must be remarked that due to the dynamic and real-time nature of the application, the purposed detection system will have to be usable on-the-go and so the usual process of calibration, such as determining std and mean results and fine-tuning weights; or attempting to prevent overfitting, is difficult to standardise. However, when tested in such conditions, we believe that the exceptionally high accuracy obtained in all metrics against the [64] dataset (General B/W Accuracy, IOU, Sørenson, low standard deviation), warrants further research and development. More standardising and analysis of the data-set will therefore be taken into consider ion in future works.

Acknowledgements

References

[1] Hemapala MU. Robots for Humanitarian Demining. In: Canbolat H, editor. Robots Operating in Hazardous Environments. IntechOpen; 2017. p. 1–21.

[2] Trevelyan J, Hamel WR, Kang SC. Robotics in Hazardous Applications. In: Siciliano B, Khatib O, editors. Springer Handbook of Robotics. Springer; 2016. p. 1521–1548.

[3] Galliott J. Military robots: Mapping the moral landscape. Routledge; 2016.

[4] Ko KH, Jang G, Park K, Kim K. GPR-Based Landmine Detection and Identification Using Multiple Features. International Journal of Antennas and Propagation. 2012;2012:7.

[5] Prado J, Marques L. Reducing false-positives in multi-sensor dataset of landmines via sensor fusion regularization. In: 2017 IEEE International Conference on Autonomous Robot Systems and Competitions, ICARSC 2017; 2017. p. 204–209.

[6] Prada PA, Rodriguez MC. Demining Dogs in Colombia - A Review of Operational Challenges, Chemical Perspectives, and Practical Implications. Science and Justice. 2016;56(4):269–277.

[7] Belkin S, Yagur-Kroll S, Kabessa Y, Korouma V, Septon T, Anati Y, et al. Remote detection of buried landmines using a bacterial sensor. Nature biotechnology. 2017;35(4):308–310.

[8] Lecun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278–2324.

[9] Koziarski M, Cyganek B. Image recognition with deep neural networks in presence of noise–dealing with and taking advantage of distortions. Integrated Computer-Aided Engineering. 2017;24(4):337–349.

[10] Molina-Cabello MA, Luque-Baena RM, López-Rubio E, Thurnhofer-Hemsi K. Vehicle type detection by ensembles of convolutional neural networks operating on super resolved images. Integrated Computer-Aided Engineering. 2018;25(4):321–333.

[11] Vera-Olmos F, Pardo E, Melero H, Malpica N. DeepEye: Deep convolutional network for pupil detection in real environments. Integrated Computer-Aided Engineering. 2019;26(1):85–95.

[12] Sanchez A, Moreno AB, Velez D, Vélez JF. Analyzing the influence of contrast in large-scale recognition of natural images. Integrated Computer-Aided Engineering. 2016;23(3):221–235.

[13] Sanchez A, Moreno AB, Velez D, Vélez JF. Analyzing the influence of contrast in large-scale recognition of natural images. Integrated Computer-Aided Engineering. 2016;23(3):221–235.

[14] Rafiei MH, Adeli H. A novel machine learning-based algorithm to detect damage in high-rise building structures. The Structural Design of Tall and Special Buildings. 2017;26(18):e1400.

[15] Cha YJ, Choi W, Büyüköztürk O. Deep Learning-Based Crack Damage Detection Using Convolutional Neural Networks. Computer-Aided Civil and Infrastructure Engineering. 2017;32(5):361–378.

[16] Yuheng S, Hao Y. Image Segmentation Algorithms Overview. arXiv e-prints. 2017 Jul;p. arXiv:1707.02051.

[17] Thoma M. A Survey of Semantic Segmentation. arXiv e-prints. 2016 Feb;p. arXiv:1602.06541.

[18] Wang XY, Wang T, Bu J. Color image segmentation using pixel wise support vector machine classification. Pattern Recognition. 2011;44(4):777 – 787.

[19] Schroff F, Criminisi A, Zisserman A. Object Class Segmentation using Random Forests. In: Proceedings of the British Machine Vision Conference. BMVA Press; 2008. p. 54.1–54.10.

[20] Yao KC, Mignotte M, Collet C, Galerne P, Burel G. Unsupervised segmentation using a self-organizing map and a noise model estimation in sonar imagery. Pattern Recognition. 2000;33(9):1575 – 1584.

[21] Rafiei MH, Khushefati WH, Demirboga R, Adeli H. Supervised Deep Restricted Boltzmann Machine for Estimation of Concrete. ACI Materials Journal. 2017;114(2).

[22] Rafiei MH, Adeli H. A novel unsupervised deep learning model for global and local health condition assessment of structures. Engineering Structures. 2018;156:598 – 607.

[23] Rafiei MH, Adeli H. Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes. Journal of Construction Engineering and Management. 2018;144(12):04018106.

[24] Ahmadlou M, Adeli H. Enhanced Probabilistic Neural Network with Local Decision Circles: A Robust Classifier. Integrated Computer-Aided Engineering. 2010 Aug;17(3):197–210.

[25] Rafiei MH, Adeli H. A New Neural Dynamic Classification Algorithm. IEEE Transactions on Neural Networks and Learning Systems. 2017;28(12):3074–3083.

[26] Wang P, Bai X. Regional parallel structure based CNN for thermal infrared face identification. Integrated Computer-Aided Engineering. 2018;25(3):247–260.

[27] Torres JF, Galicia A, Troncoso A, Martínez-Álvarez F. A scalable approach based on deep learning for big data time series forecasting. Integrated Computer-Aided Engineering. 2018;25(4):335–348.

[28] Yang T, Cappelle C, Ruichek Y, El Bagdouri M. Multi-object tracking with discriminant correlation filter based deep learning tracker. Integrated Computer-Aided Engineering. 2019;26(3):273–284.

[29] Vera-Olmos F, Pardo E, Melero H, Malpica N. DeepEye: Deep convolutional network for pupil detection in real environments. Integrated Computer-Aided Engineering. 2019;26(1):85–95.

[30] Oh B, Glisic B, Kim Y, Park H. Convolutional neural network-based wind-induced response estimation model for tall buildings. Computer-Aided Civil and Infrastructure Engineering. 2019;34(10):843–858.

[31] Shen J, Xiong X, Xue Z, Bian Y. A convolutional neural-network-based pedestrian counting model for various crowded scenes. Computer-Aided Civil and Infrastructure Engineering. 2019;34(10):897–914.

[32] Maeda K, Takahashi S, Ogawa T, Haseyama M. Convolutional sparse coding-based deep random vector functional link network for distress classification of road structures. Computer-Aided Civil and Infrastructure Engineering. 2019;34(8):654–676.

[33] Wu RT, Singla A, Jahanshahi MR, Bertino E, Ko BJ, Verma D. Pruning deep convolutional neural networks for efficient edge computing in condition assessment of infrastructures. Computer-Aided Civil and Infrastructure Engineering. 2019;34(9):774–789.

[34] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical Image Analysis. 2017;42:60 – 88.

[35] Schwartz G, Nishino K. Recognizing Material Properties from Images. arXiv e-prints. 2018 Aug;p. arXiv:1801.03127.

[36] Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. In: Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1. NIPS'12. Curran; 2012. p. 1097–1105.

[37] Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv e-prints. 2014 Sep;p. arXiv:1409.1556.

[38] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going Deeper with Convolutions. arXiv e-prints. 2014 Sep;p. arXiv:1409.4842.

[39] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv e-prints. 2015 Dec;p. arXiv:1512.03385.

[40] Al-Qizwini M, Barjasteh I, Al-Qassab H, Radha H. Deep learning algorithm for autonomous driving using GoogLeNet. In: 2017 IEEE Intelligent Vehicles Symposium (IV); 2017. p. 89–96.

[41] Tian Y, Pei K, Jana S, Ray B. DeepTest: Automated Testing of Deep-neural-network-driven Autonomous Cars. In: Proceedings of the 40th International Conference on Software Engineering. ICSE '18. ACM; 2018. p. 303–314.

[42] Sallab AE, Abdou M, Perot E, Yogamani S. Deep reinforcement learning framework for autonomous driving. Electronic Imaging. 2017;2017(19):70–76.

[43] Shen D, Wu G, Suk HI. Deep Learning in Medical Image Analysis. Annual Review of Biomedical Engineering. 2017;19(1):221–248.

[44] Ortiz A, Munilla J, Górriz JM, Ramírez J. Ensembles of Deep Learning Architectures for the Early Diagnosis of the Alzheimer's Disease. International Journal of Neural Systems. 2016;26(07):1650025.

[45] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. arXiv e-prints. 2015 May;p. arXiv:1505.04597.

[46] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. Computers in Biology and Medicine. 2018;100:270 – 278.

[47] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H, Subha DP. Automated EEG-based screening of depression using deep convolutional neural network. Computer Methods and Programs in Biomedicine. 2018;161:103 – 113.

[48] Acharya UR, Hagiwara Y, Adeli H. Automated seizure prediction. Epilepsy & Behavior. 2018;88:251 – 261.

[49] Sankari Z, Adeli H. Probabilistic neural networks for diagnosis of Alzheimer's disease using conventional and wavelet coherence. Journal of Neuroscience Methods. 2011;197(1):165 – 170.

[50] Acharya UR, Oh SL, Hagiwara Y, Tan JH, Adeli H, Subha DP. Automated EEG-based screening of depression using deep convolutional neural network. Computer Methods and Programs in Biomedicine. 2018;161:103 – 113.

[51] Takahashi K, Preetz H, Igel J. Soil properties and performance of landmine detection by metal detector and ground-penetrating radar - Soil characterisation and its verification by a field test. Journal of Applied Geophysics. 2011;73(4):368–377.

[52] Byrnes J. Unexploded Ordnance Detection and Mitigation. Springer; 2008.

[53] Bhope PC, Bhalchandra AS. Various Landmine Detection Techniques: A Review. International Journal of Innovative Research in Science, Engineering and Technology. 2015;4(6):771–775.

[54] Lopez P, Vilarino DL, Cabello D. CNN-based 3D thermal modeling of the soil for antipersonnel mine detection. In: Tetzlaff R, editor. Proceedings of the 2002 7th IEEE International Workshop on Cellular Neural Networks and Their Applications. World Scientific; 2002. p. 307–314.

[55] Stone K, Keller JM. Convolutional neural network approach for buried target recognition in FL-LWIR imagery. In: Bishop SS, Isaacs JC, editors. Prodeedings of SPIE Defense + Security: Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIX. SPIE; 2014. p. 305 – 319.

[56] Yilmaz C, Kahraman H, Söyler S. Passive Mine Detection and Classification Method Based on Hybrid Model. IEEE Access. 2018;6:47870 – 47888.

[57] Kafedziski V, Pecov S, Tanevski D. Detection and Classification of Land Mines from Ground Penetrating Radar Data Using Faster R-CNN. In: 2018 26th Telecommunications Forum (TELFOR). IEEE; 2018. .

[58] Lameri S, Lombardi F, Bestagini P, Lualdi M, Tubaro S. Landmine detection from GPR data using convolutional neural networks. In: 2017 25th European Signal Processing Conference (EUSIPCO). IEEE; 2017. p. 508–512.

[59] Dzieciuch I, Gebhardt D, Barngrover C, Parikh K. Nonlinear Convolutional Neural Network for Automatic Detection of Mine-Like Objects in Sonar Imagery. In: In V, Longhini P, Palacios A, editors. Proceedings of the 4th International Conference on Applications in Nonlinear Dynamics (ICAND 2016). Springer; 2017. p. 309–314.

[60] Caraffini F, Florez-Lozano J, Gongora M, Parra C. Multi-agent robotic system dataset. figshare; 2019. Doi:10.21253/DMU.c.4555424.v2. https://doi.org/10.21253/DMU.c.4555424.v2.

[61] Florez-Lozano J, Caraffini F, Gongora M, Parra C. CoD2M-MAPS Source Code. figshare; 2019. Doi:10.21253/DMU.9630779.v1. https://doi.org/10.21253/DMU.9630779.v1.

[62] Caraffini F, Gongora M. Collaborative methodology for enhancing sustainability in rural communities and the use of land. figshare; 2019. Doi:10.21253/DMU.8483960. https://doi.org/10.21253/DMU.8483960.

[63] Florez-Lozano J, Caraffini F, Gongora M, Parra C. Stitched images & masks - August 2017. figshare; 2019. Doi:10.21253/DMU.8314520.v1. https://doi.org/10.21253/DMU.8314520.v1.

[64] Silva JS, Guerra IFL, Bioucas-Dias J, Gasche T. Landmine Detection Using Multispectral Images. IEEE Sensors Journal. 2019;19(20):9341–9351.

[65] Badrinarayanan V, Kendall A, Cipolla R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2017;39(12):2481–2495.

[66] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, et al. Recent advances in convolutional neural networks. Pattern Recognition. 2018;77:354 – 377.

[67] Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y. What is the best multi-stage architecture for object recognition? In: 2009 IEEE 12th International Conference on Computer Vision. IEEE; 2009. p. 2146–2153.

[68] Chen M, Shi X, Zhang Y, Wu D, Guizani M. Deep Features Learning for Medical Image Analysis with Convolutional Au-

toencoder Neural Network. IEEE Transactions on Big Data. 2018;p. 1.

[69] Team TTD, Al-Rfou R, Alain G, Almahairi A, Angermueller C, Bahdanau D, et al. Theano: A Python framework for fast computation of mathematical expressions. arXiv preprint arXiv:160502688. 2016;.

[70] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in Python. Journal of machine learning research. 2011;12:2825–2830.

[71] Garreta R, Moncecchi G. Learning scikit-learn: machine learning in python. Packt Publishing; 2013.

[72] Glumov NI, Kolomiyetz EI, Sergeyev VV. Detection of objects on the image using a sliding window mode. Optics & Laser Technology. 1995;27(4):241 – 249.

[73] Jaccard P. The Distribution of the flora in the alpine zone.1. New Phytologist. 1912;11(2):37–50.

[74] Sørensen TA. A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. Biologiske skrifter. 1948;5:1–34.

[75] Le H, Vicente TFY, Nguyen V, Hoai M, Samaras D. A+D Net: Training a Shadow Detector with Adversarial Shadow Attenuation. In: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, editors. Computer Vision – ECCV 2018. Springer; 2018. p. 680–696.