Using Manifold Embedding for Automatic Threat Detection: An Alternative Machine Learning Approach

Roberta Piroddi[1]
R.Piroddi@liverpool.ac.uk
Elias Griffith[1]
E.Griffith@liverpool.ac.uk
Yannis Goulermas[2]
J.Y.Goulermas@liverpool.ac.uk
Simon Maskell[1]
S.Maskell@liverpool.ac.uk
Jason Ralph[1]
Jfralph@liverpool.ac.uk

[1] Department of Electrical Engineering and Electronics
EEE Building A,
Brownlow Hill,
University of Liverpool
Liverpool, UK

[2] Department of Computer Science
Ashton Building,
The Quadrangle,
University of Liverpool,
Liverpool, UK

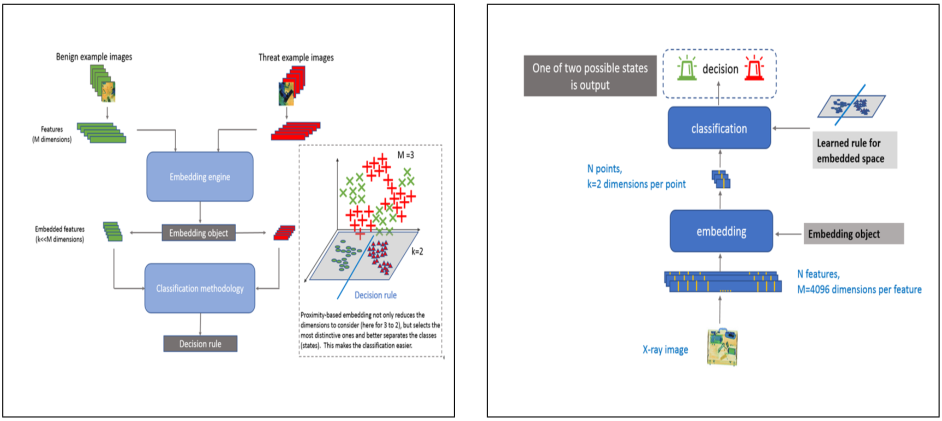
Abstract

This paper presents an approach to automatic threat detection in X-ray imagery based on the combination of manifold embedding and a specialized classification strategy to identify a wide range of firearm threats in imagery captured in operational settings and in challenging simulated settings which closely reflected operational conditions. The results show that this approach is successful in reducing the amount of data needed to produce a flexible classification system by two orders of magnitude if compared to most state-of-the-art deep learning systems used for object recognition. The work also offers useful indications to the amount of data required for such an alternative machine learning approach. It also shows an enhanced algorithmic architecture for the fusion of very different classifiers, so that a generic and flexible system may be adapted to varied operational environments.

1 Introduction

In recent years machine learning, especially in the embodiment of deep learning, has been responsible for great feats of automation in tasks that previously had been considered only in the realm of human processing capabilities [18]. The object recognition task on unstructured and large image data-sets has been the most successfully devolved to machine analysis, in terms of volume, speed and accuracy of the recognition [14, 31].

It is therefore sensible to apply a machine-learning approach to the task of firearm threat detection and recognition in security screening [13, 20]. The application of automation, let alone machine learning, to such security applications, is particularly challenging [12, 13]. Effective systems need to cope with the operational needs of detecting both full weapons and



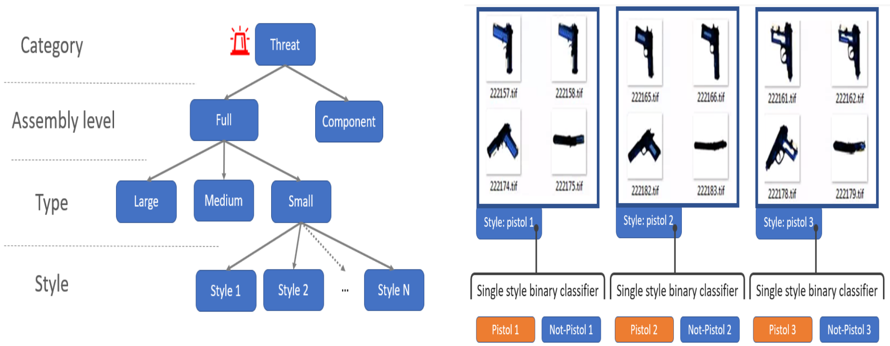
(a) (b)

Figure 1: (a) Training phase of our MOPE-SVN system (b) Analysis phase where a scanned image is assigned one of two labels.

components, in environments that span from the security screening of travellers luggage, to the border checks of goods transported in pallets and containers. X-ray sensors in use for security screening applications vary substantially in the characteristics of the imagery that they are able to provide as a function of the environment where they are deployed [7, 28], settings decided by single operators, and colour depth [6, 21]. In contrast to those employed in successful machine learning, data-sets available to train security screening applications are typically unbalanced: the vast majority of examples are of benign images, with few examples of images containing threats [1]. Finally, the system performance accuracy that may deemed to be acceptable for such sensitive applications needs to be extremely high and not just when considering the detection of threats, but also with regards to the excessive presence of false positive alerts that may severely dent the trust of the operator in the system, in an environment characterized by high time pressure [5].

Most other studies on automated threat detection for security screening which consider specific subsets of challenges [22], for example the development of a classifier for a specific class of threats [1, 4, 25] in a restricted [15, 16, 19] and very controlled environment [11, 27, 28]. These works show that for deep learning to be highly effective in this application it requires a number of examples of target object that is unusual to obtain. The method developed here therefore considers an alternative machine learning approach, which is able to learn from fewer examples to deliver a flexible and accurate system with realistic quantities of target imagery. This approach uses manifold embedding [29] to discover the most distinctive sets of features to describe a target object and it combines it with a specialized classification strategy to maximize the potentiality of the embedding technique to distill salient information. Although manifold embedding is well-known mainly for dimensionality reduction, we applied it as a saliency detector using a technique, MOPE, that has been shown to generalize well from small data-sets [23]. Around this core concept (see Figure 1), this approach builds a full system and processing pipeline for concrete, real or realistic and extremely challenging operational boundaries of use.

This paper is structured as follows. Section 2 discusses the architecture of a system to answer the requirements of adaptability and flexibility of target and environment. Section 3 describes the experimental set-up we used to train and test the system. Section 4 provides



(a) (b)

Figure 2: (a) Object hierarchy for the category with the label of threat. (b) Example of style of objects for a particular type of threat and the concept of binary classifier developed to detect each style.

the performance characteristics of the proposed system. Wider implications of this work are discussed in Section 5. Conclusions follow in Section 6.

2 Method

This machine learning multiple-stage architecture has the ultimate goal to recognize between two possible states of the sensed world: one state is for a given X-ray image to contain a firearm (or part of it) and the second state is for a given image to contain only benign objects [9].

The system will therefore be composed of two main subsystems to train the classifier and then to apply it to analyze an image, see Figure 1. In an operational setting, when a new image is input into the system, according to the learned classification rules, it gives the new image a label, either benign or threat [3]. The characteristics of the system depend on the machine learning approach and the classification strategy used.

2.1 Machine learning analytic core

The core classifier is a Support Vector Machine (SVM) [9], which requires a choice of features [10, 30]. Manifold embedding, irrespective of features employed, would help to sift salient information [10, 26], especially considering different poses of the same object [24, 32]. Multi-Output Proximity Embedding (MOPE) [23] has shown to generalize well from a relatively small sample of training exemplars [8]. In the recognition system proposed, see Figure 1, MOPE sits at the interface between the feature extraction and classification. The input to the classifier is an embedded set of features that are well separated into clusters corresponding to the two possible states. A linear kernel for the SVM is sufficient to find an efficient classification rule, shown diagrammatically in the inset of Figure 1.

In the testing or analysis phase, which is when this system is operationally used, an image never seen by the system is presented in input. The new image is classified either as an acceptable (benign) or an anomalous (threat) state. A simplified view of the operational

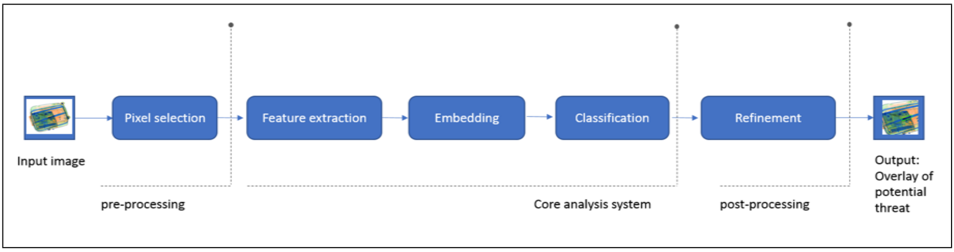


Figure 3: Analysis pipeline employed to apply one classifier to a test image.

system is shown in Figure 1.(b). To ensure that the capability of embedding for saliency extraction is maximized, the classification engine is coupled with a single classifier development strategy where highly specialized binary classifiers are developed for specific styles of threats, as shown in Figure 2.(b). This strategy hinges on having a hierarchy of objects to detect, as shown in Figure 2.(a).

2.2 Analysis pipeline

The system's unit of analysis is a processing pipeline that uses a single classifier (as shown in Figure 2.(b)) and applies it to analyze a test image, as in Figure 3.

Pre-processing is performed using color (some X-ray scanners provide dual color imagery, where the color depends on the energy band), where possible, and intensity information. With color imagery, highly informative colour signature can be associated to potential threats. The core analysis module extracts Discrete Cosine Transform (DCT) [2, 9] features on square patches of the image, obtaining 4096 features for each window around a pixel selected for processing. They are then transformed using the MOPE to 2. Each pixel is classified according to its embedded feature values using a linear SVM.

Post-processing is performed on the basis of the color signature of contiguous areas of pixels classified as threat. Using the statistics of the size of threats relative sensor resolution, it is possible to further exclude segments.

A single classifier is specific to the environment, because the type of clutter and benign objects are different in luggage item, a pallet or a container [7, 30], so this pipeline needs to be employed for each environment with its specific single binary classifiers.

2.3 Classifier bank and fusion

In this approach, the categories of benign and threat are summaries of specific categories of threats (as in Figure 2.(a)). Classifier fusion is required to get to the final distinction between threat and benign. The classifiers are resolution specific, so each classifier for each type of firearm in each environment has an associated resolution. The multi-resolution part of the processing retrieves the information about the link between classifier and resolution.

The system employs Bayesian average classifier fusion [17]. The probability associated to each label is based on color segmentation, considering the overlap between segments which present suspicious characteristics and those classified as threats. Where the sensor is a grey-level sensor, intensity is used instead. The fusion pipeline is shown in Figure 4.

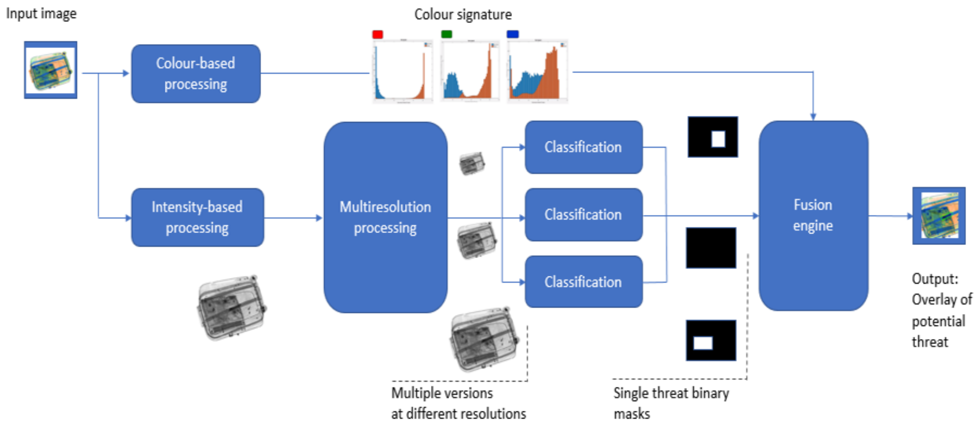


Figure 4: Whole processing pipeline to test multiple classifiers on one candidate test image.

Each environment needs its own analysis tool with classifier banks developed to find threats and components in the clutter, and sensor resolution characteristics, that are associated to the imaged environment.

Figure 4 shows how the processing components are connected together to support the need to cater for different threats and sensing environment. At input, the program settings indicate the environment and the kind of sensor (e.g. resolution as a function of environment and colour intensity characteristics). If it is applicable, the dual colour component is split from the intensity and processed to extract and classify colour regions. A set of multi-resolution versions of the input images are generated and presented as input to embedding modules. This is required because the embedding and following classifiers are resolution dependent and to classify different styles of threats and components, each of the specialized classifiers is linked to a specific resolution of the image in a given environment. Each classifier outputs a binary image containing bounding boxes of areas of the image of potential threat. The multiple binary masks, relative to different resolutions and threat types, are collected and re-sized to a common resolution. There may be bounding boxes that overlap and in order to provide a clear label, the bounding boxes are given as inputs to a fusion engine. This engine uses the colour signatures and links them to the areas classified as potential threat. This provides a probability estimation of a region contained in a bounding box being a certain threat. The probabilities are fused together using Bayesian averaging so that the label with the highest probability is output with a colour coded overlay showing a bounding box relative to the label with the highest probability.

3 Experimental set-up

The UK Home Office provided the X-ray data-set as part of a funding competition to support automatic threat detection research. Experiments covered three different environments: luggage, as baggage (airport hand baggage type of environment) and fast parcels (hold-type baggage), and pallets and containers. The sensors characteristics, the image resolution and the image quality as well as clutter were different for each kind of environment [21]. For

TYPE: PISTOL	STYLE 1		STYLE 2		STYLE 3	
	Generalized classifier	Specialized classifier	Generalized classifier	Specialized classifier	Generalized classifier	Specialized classifier
True positives	92%	92%	92%	96%	92%	90%
False Positives	178%	246%	178%	132%	178%	50%

Figure 5: Performance comparison between the strategy of training a classifier on a generic object type (in this case a pistol) and training a classifier specialized on a particular style of the object.

each environment, the task was to detect full firearms and components.

Training. Each classifier was trained using 10^4 training samples, of these 70% were of non-target examples (negatives or benign). To build positive (threat) examples, reference images of the full threats were superimposed to background clutter samples where available. Reference images of components were unavailable in general. Where references were unavailable, the testing set was split and a subset was used segmented to use for training. For full-weapons in luggage the data-set consisted of 5×10^5 operational benign images and tens of reference images for around 10 firearm types in 52 out-of-plane 3D poses, taken with consistent and high resolution dual color airport scanners. For pallets and containers, reference images were not available. Container sensors provided grey-level imagery. The data-set contained around 80 benign images for pallets and 7 for containers.

Testing. The experimental plan covered separately the three different environments, with classifiers trained for full firearms for the three environment. Weapon component classifiers were developed for the luggage environment only. It was impossible to develop component classifiers for the larger environments, because of (a) the low resolution of the images making the components too small to be processed and (b) the contrast characteristics of the imagery which meant that the layers of clutter were able to obscure the smallest components. The experimental plan also contained several tasks, each one related to finding a particular kind of threat (full, component). Full threats are subdivided into: small sizes (subdivided into styles), medium sizes, and large sizes. The component classifiers were: barrel, bolt, trigger, trigger assembly and spring. The experiments considered also the evaluation of the classification strategy that designs numerous specialized classifiers.

4 Results

The first set of experimental results concern the evaluation of the classification strategy employed to maximize the potential of the embedding technique to abstract from the feature selection and still extract the more distinctive information from a set of features, as presented in Section 2.1. The performance of this strategy compared to the use of a generalized classifier that is trained on a general category of firearm is presented in Figure 5. It is possible to increase the accuracy of the classification in terms of both true and false positives, sometimes seemingly with some trade off between false and true positives.

		LUGGAGE		PALLET		CONTAINER	
benign / threat		Predicted		Predicted		Predicted	
		Positive	Negative	Positive	Negative	Positive	Negative
Actual	Positive	97.3%	2.7%	100%	0%	0%	100%
	Negative	52%	48%	100%	0%	100%	0%

Figure 6: Classification results for a choice of labels between threat and benign for the three environments of luggage, pallet and container and for classifiers trained on full objects.

			Predicted			
L M S B			Large	Medium	Small	Benign
Actual	Large	L	14%	83%	3%	0%
	Medium	M	14%	86%	0%	0%
	Small	S	6%	89%	1%	4%
	Benign	B	8%	30%	14%	48%

(a)

			Predicted			
L M S B			Large	Medium	Small	Benign
Actual	Large	L	50%	0%	50%	0%
	Medium	M	5%	28%	67%	0%
	Small	S	10%	3%	83%	4%
	Benign	B	0.5%	0.5%	5%	94%

(b)

Figure 7: (a) MOPE-SVM performance after partial fusion distinguishing between 4 labels (3 threats and 1 benign) and (b) CNN performance trained on the same labels and and tested on the same data sub-set.

The performance of the combined classification between threat and benign, after processing described in Section 2.3 when all filters trained on full weapons are considered is shown in Figure 6. The classification is not meaningful for pallets and containers, possibly because of insufficient number and variability of benign images. The lack of reference images from sensors specific to such environments may have also played a role.

For the luggage environment, Figure 7.(a) shows the performance of the system in detecting specific types of full weapon: large weapons (e.g. assault rifles), medium sized weapons (e.g. sub-machine guns), and small weapons (e.g. pistols). It was possible to train a Convolutional Neural Network (CNN) using some limited transfer learning on the same labels and then test it on a set of very well-behaved test images. The CNN performance is given in 7.(b), as reference point to a theoretical best-case scenario.

Classifiers trained to recognize components were only evaluated in the luggage environment, with two kinds of experiments. The first was using test images with known components. The performance is shown in Figure 8.(a).

A second kind of experiment was performed on images containing full weapons or benign images. This experiment allowed investigating: (a) whether classifiers that were developed to find components may find parts of full objects, and (b) how many benign objects were classified as possible weapon parts. The results of this experiment are reported in Figure 8.(b).

The classifier for the barrel seemed to include features that made it the most recognizable part. Parts as spring even though they did look recognizable, contained so many frequencies in the coils and different variations and non-rigid transformations that were not easily encapsulated in only one non-invariant classifier. The approach to develop classifiers from

Actual \ Predicted	bolt	barrel	trigger assembly	trigger	spring
bolt	12%	95%	17%	0%	46
barrel	16%	98%	40%	0%	83
trigger assembly	28%	98%	44%	0%	89
trigger	16%	50%	46%	0%	54
spring	8%	85%	10%	0%	26%

(a)

Actual \ Predicted	bolt	barrel	trigger assembly	trigger	spring
full weapon	28%	96%	100%	0	56%
benign	6%	92%	17%	0	44%

(b)

Figure 8: Confusion matrix for (a) a set of 5 categories of threat components for which a classifier has been built in the luggage environment and (b) for component classifiers tested on full weapons and benign images.

cut-out of images (as the trigger) did not work, because this approach used global and not part-models. For this reason, the trigger assembly, a part with similar shape and size to a full threat achieved good results on full threat images, while the trigger did not.

Finally, to give an idea of the intermediate and final outputs and the quality of the results obtained using the full analysis pipeline, in Figure 9 shows the processing steps that led to the final classification label for a few sample images.

5 Discussion

Classification performance is linked to the amount and the quality of examples in the training phase. The ability of the training data-set to cover a large part of the space of possible variations of operational scenarios is crucial to all machine learning methods. The classifiers evaluated in this project have been trained using $10,000$ samples. Experiments have shown that for each $1,000$ extra samples there was on average an accuracy increase of 0.25%. Deep learning systems are trained using millions of samples. Achieving such a large number of images in this project was difficult, with very few examples of background clutter in pallets or containers, and few or absent reference images of threats.

The performance characteristics show that it is possible to increase the accuracy of the classification by employing very specialized classifiers. This approach requires to have reference images to support the specialization in terms of object styles. It also requires increased sophistication in terms of the algorithmic architecture that supports the classifier fusion. This also has implications in the computational complexity.

The approach here proposed is successful in terms of true positives identified, with performance comparable on this metric with deep learning, using two orders of magnitude fewer samples for the training. Still the MOPE-SVM approach requires sufficient numbers of benign examples and reference examples of threats generated by sensors that are environment specific. For the environments, like pallets and containers, where such reference images and fewer benign examples were available, the performance showed that it is challenging to transfer characteristics and statistical signatures from one sensor and environment to another space. This is particularly important in case of transfer from colour to grey-level sensors. This is line with the finding of previous studies [6, 7, 10].

Further research is needed to establish the optimal balance of parameters in terms of

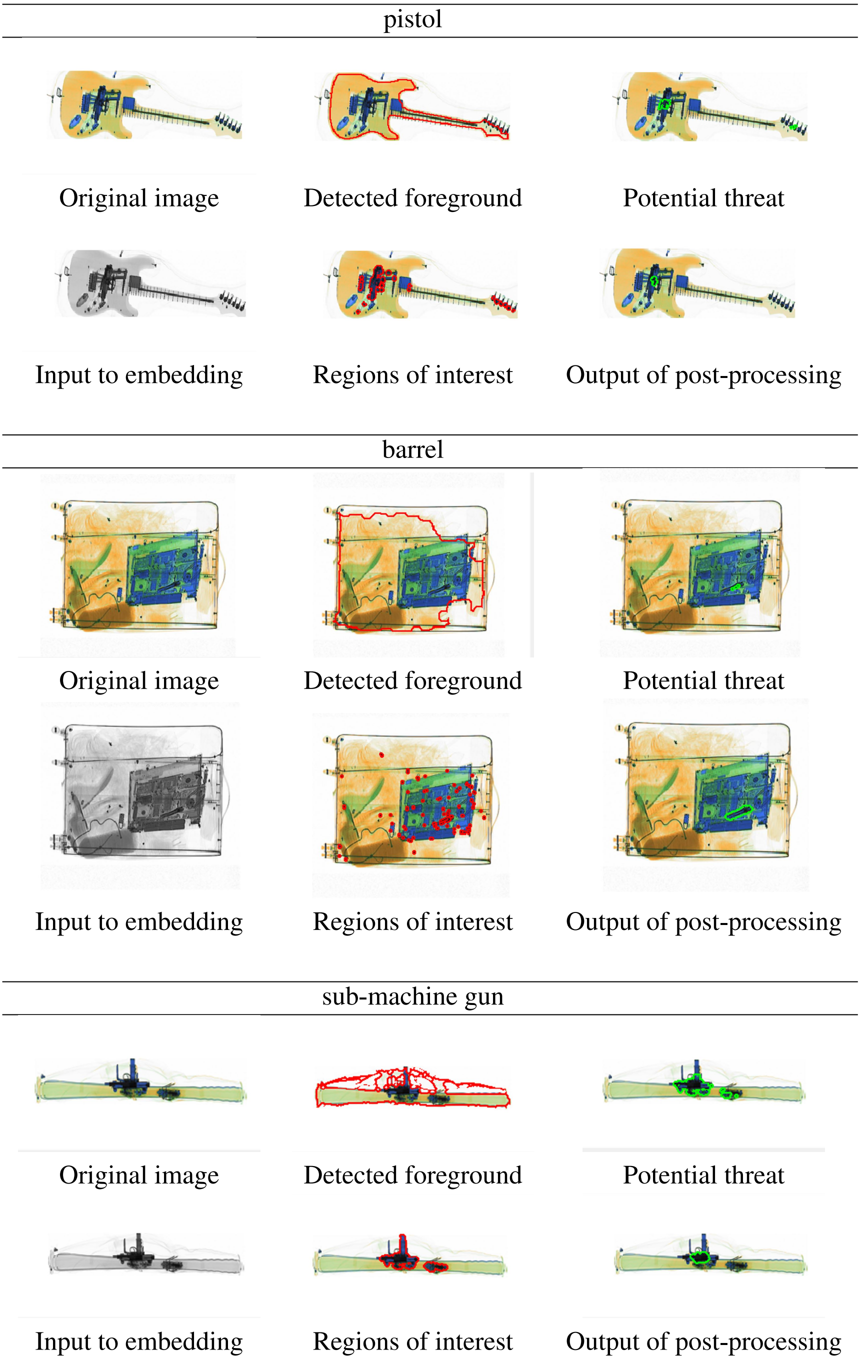


Figure 9: Intermediate analysis results of MOPE-SVM analysis pipeline in images of luggage environment.

samples and specialization required to support an operationally useful result. This is especially the case for firearm component detection. Another area where research is needed is that of the link between classification performance and sensor characteristics, in order to find the optimal method to tune the two. This is likely to be a fast moving area as new types of sensing technology emerge.

6 Conclusions

This paper presented a multistage analysis and classification system to perform automatic threat detection in variable security screening environments. The system contained specific algorithms to deal flexibly with multiple categories of threats and multiple screening environments ranging from hand-held luggage at airports to container in good transportation scenarios. The analysis of the performance characteristics of the proposed approach showed that combining an embedding method that selected discriminative information with a classification strategy that maximized the opportunity to locate salient information was a viable method to apply machine learning for threat detection when the quantity of data was small. This approach used roughly two orders of magnitude less training data than deep learning and achieved accuracy comparable to state-of-the-art Convolutional Neural Networks for true positives. The results also confirmed that, like expressed elsewhere in the literature [22], feature transfer from one sensor to another was challenging. Therefore systems needed to be developed for specific environments. More research is required to develop efficient methods tailored to deployment environments. As of state-of-the-art X-ray sensors for screening, more research is needed to establish the parameters that render classification systems more effective for each environment specificity and presentation.

Acknowledgement

The authors would like to thank the UK Home Office for partially funding this work, under the UK Home Office grant: OSCT CONTEST S&T BORD-17-255. The authors wish to acknowledge Mr Richard Jackson, Defence Science and Technology Laboratory, for liaising with the Home Office, Office for Security and Counter Terrorism (OSCT), offering steering of this project and helpful discussion, and reviewing outputs. Views contained within this paper are not necessarily those of the UK Home Office.

References

[1] S. Akcay, M. E. Kundegorski, C. G. Willcocks, and T. P. Breckon. Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. *IEEE Transactions on Information Forensics and Security*, 13(9):2203–2215, 2018.

[2] E. Alickovic, J. Kevric, and A. Subasi. Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction. *Biomedical Signal Processing and Control*, 39:94–102, 2018.

[3] J. Atanbori, W. Duan, J. Murray, K. Appiah, and P. Dickinson. A computer vision approach to classification of birds in flight from video sequences. In *Proceedings of British Machine Vision Conference*, pages 3.1–3.9. British Machine Vision Association, 2015.

[4] M. Bastan, W. Byeon, and T. M. Breuel. Object recognition in multi-view dual energy x-ray images. In *Proceedings of British Machine Vision Conference*, page 11. British Machine Vision Association, 2013.

[5] A. Bolfing, T. Halbherr, and A. Schwaninger. How image based factors and human factors contribute to threat detection performance in x-ray aviation security screening. In *symposium of the Austrian HCI and Usability Engineering Group*, pages 419–438. Springer, 2008.

[6] K. Brumbaugh, C. Royse, C. Gregory, K. Roe, J.A. Greenberg, and S.O. Diallo. Material classification using convolution neural network (cnn) for x-ray based coded aperture diffraction system. In *Anomaly Detection and Imaging with X-Rays (ADIX) IV*, volume 10999, page 109990B. International Society for Optics and Photonics, 2019.

[7] M. Caldwell, M. Ransley, T. W. Rogers, and L. D. Griffin. Transferring x-ray based automated threat detection between scanners with different energies and resolution. In *Counterterrorism, Crime Fighting, Forensics, and Surveillance Technologies*, volume 10441, page 104410F. International Society for Optics and Photonics, 2017.

[8] Y. Chi, E. J. Griffith, J. Y. Goulermas, and J. F. Ralph. Binary data embedding framework for multiclass classification. *IEEE Transactions on Human-Machine Systems*, 45 (4):453–464, 2015.

[9] R. B. Fisher, T. P. Breckon, K. Dawson-Howe, A. Fitzgibbon, C. Robertson, E. Trucco, and C. K.I. Williams. *Dictionary of computer vision and image processing*. John Wiley & Sons, 2013.

[10] G. Flitton, A. Mouton, and T. P. Breckon. Object classification in 3d baggage security computed tomography imagery using visual codebooks. *Pattern Recognition*, 48(8): 2489–2499, 2015.

[11] L. D. Griffin, M. Caldwell, J. T.A. Andrews, and H. Bohler. âĂIJunexpected item in the bagging areaâĂİ: Anomaly detection in x-ray security images. *IEEE Transactions on Information Forensics and Security*, 14(6):1539–1553, 2018.

[12] T. Harvey. Aviation security x-ray detection challenges. In *Anomaly Detection and Imaging with X-Rays (ADIX)*, volume 9847, page 984702. International Society for Optics and Photonics, 2016.

[13] N. Hättenschwiler, Y. Sterchi, M. Mendes, and A. Schwaninger. Automation in airport security x-ray screening of cabin baggage: Examining benefits and possible implementations of automated explosives detection. *Applied ergonomics*, 72:58–68, 2018.

[14] G. Huang, G-B. Huang, S. Song, and K. You. Trends in extreme learning machines: A review. *Neural Networks*, 61:32–48, 2015.

[15] N. Jaccard, T. W. Rogers, E. J. Morton, and L. D. Griffin. Detection of concealed cars in complex cargo x-ray imagery using deep learning. *Journal of X-ray Science and Technology*, 25(3):323–339, 2017.

[16] D. K. Jain. An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery. *Pattern Recognition Letters*, 120:112–119, 2019.

[17] H-C. Kim and Z. Ghahramani. Bayesian classifier combination. In *Artificial Intelligence and Statistics*, pages 619–627, 2012.

[18] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436, 2015.

[19] K. J. Liang, G. Heilmann, C. Gregory, S. O. Diallo, D. Carlson, G. P. Spell, J. B. Sigman, K. Roe, and L. Carin. Automatic threat recognition of prohibited items at aviation checkpoint with x-ray imaging: A deep learning approach. In *Anomaly Detection and Imaging with X-Rays (ADIX) III*, volume 10632, page 1063203. International Society for Optics and Photonics, 2018.

[20] D. Mery. *Computer Vision for X-Ray Testing*. Springer, 2015.

[21] A. Mouton and T. P. Breckon. Materials-based 3d segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening. *Pattern Recognition*, 48(6):1961–1978, 2015.

[22] A. Mouton and T. P. Breckon. A review of automated image understanding within 3d baggage computed tomography security screening. *Journal of X-ray science and technology*, 23(5):531–555, 2015.

[23] T. Mu, J. Y. Goulermas, J. Tsujii, and S. Ananiadou. Proximity-based frameworks for generating embeddings from multi-output data. *IEEE transactions on pattern analysis and machine intelligence*, 34(11):2216–2232, 2012.

[24] F. Nie, W. Zhu, and X. Li. Unsupervised feature selection with structured graph optimization. In *Thirtieth AAAI conference on artificial intelligence*, 2016.

[25] C. Paulus, V. Moulin, D. Perion, P. Radisson, and L. Verger. Multi-energy x-ray detectors to improve air-cargo security. In *Anomaly Detection and Imaging with X-Rays*, volume 10187, page 101870I. International Society for Optics and Photonics, 2017.

[26] R. Piroddi, J.Y. Goulermas, S. Maskell, and J.F. Ralph. Comparing interrelationships between features and embedding methods for multiple-view fusion. In *2018 21st International Conference on Information Fusion*, pages 1–5. IEEE, 2018.

[27] V. Riffo and D. Mery. Automated detection of threat objects using adapted implicit shape model. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 46(4): 472–482, 2015.

[28] T. W. Rogers, N. Jaccard, E. J. Morton, and L. D. Griffin. Automated x-ray image analysis for cargo security: Critical review and future promise. *Journal of X-ray science and technology*, 25(1):33–56, 2017.

[29] D. Tao, J. Cheng, M. Song, and X. Lin. Manifold ranking-based matrix factorization for saliency detection. *IEEE transactions on neural networks and learning systems*, 27 (6):1122–1134, 2015.

[30] D. Turcsany, A. Mouton, and T. P. Breckon. Improving feature-based object recognition for x-ray baggage security screening using primed visualwords. In *2013 IEEE International Conference on Industrial Technology (ICIT)*, pages 1140–1145. IEEE, 2013.

[31] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis. Deep learning for computer vision: A brief review. *Computational intelligence and neuroscience*, 2018, 2018.

[32] Y. Zhao, X. You, S. Yu, C. Xu, W. Yuan, X-Y. Jing, T. Zhang, and D. Tao. Multi-view manifold learning with locality alignment. *Pattern Recognition*, 78:154–166, 2018.