Joint Sub-component Level Segmentation and Classification for Anomaly Detection within Dual-Energy X-Ray Security Imagery

Neelanjan Bhowmik[1], Toby P. Breckon[1,2]
Department of {Computer Science[1] | Engineering[2]}, Durham University, UK

***Abstract*—X-ray baggage security screening is in widespread use and crucial to maintaining transport security for threat/anomaly detection tasks. The automatic detection of anomaly, which is concealed within cluttered and complex electronics/electrical items, using 2D X-ray imagery is of primary interest in recent years. We address this task by introducing joint object sub-component level segmentation and classification strategy using deep Convolution Neural Network architecture. The performance is evaluated over a dataset of cluttered X-ray baggage security imagery, consisting of consumer electrical and electronics items using variants of dual-energy X-ray imagery (pseudo-colour, high, low, and effective-Z). The proposed joint sub-component level segmentation and classification approach achieve $\sim 99\%$ true positive and $\sim 5\%$ false positive for anomaly detection task.**

***Index Terms*—X-ray imagery, superpixel, deep convolutional neural network, anomaly detection, classification.**

I. Introduction

With the increasing volume of traffic, we need to ensure an efficient system for aviation securing capable of addressing the evolving threat landscape that emanates from broader global geopolitical events. Currently multiple-view X-ray baggage security screening is widely used to maintain aviation and transport including the screening of electronics/electrical items. To address the future challenges of increasing volumes and complexities, the recent focus on the use of automated screening approaches is of particular interest. This includes the potential for automatic anomaly/threat detection as a methodology for concealment detection within complex electronics and electrical items screened using low-cost, 2D X-ray imagery. Passenger baggage is currently inspected manually using dual-energy multiple-view X-ray imaging. The threat concealment can also be very subtle and very well hidden (Figure 1A) challenging for a human operator to identify.

Early work on automated threat detection within X-ray security images is based on hand-crafted features [1], [2] (Bag-of-Visual-Words), which is applied together with a classifier such as a Support Vector Machine. More recent work [3], [4], that specifically leverage recent advances in Convolutional Neural Networks (CNN) deep learning architectures [5], [6], have now been shown to outperform earlier approaches in terms of true positive detection, false alarm rate and the range of objects that can be detected in a side by side comparison. By contrast, CNN approach of [3] operates at scan rate ($<$1 sec. per image), with higher accuracy and targets probabilistic item localisation within each X-ray view. The work of [7] considers a unique feature representation as a critical component for detection within cluttered X-ray imagery for anomaly detection. In the works of [8], [9], semi-supervised anomaly detection strategies are proposed based on high reconstruction errors produced by a generator network adversarially trained on benign X-ray imagery only.

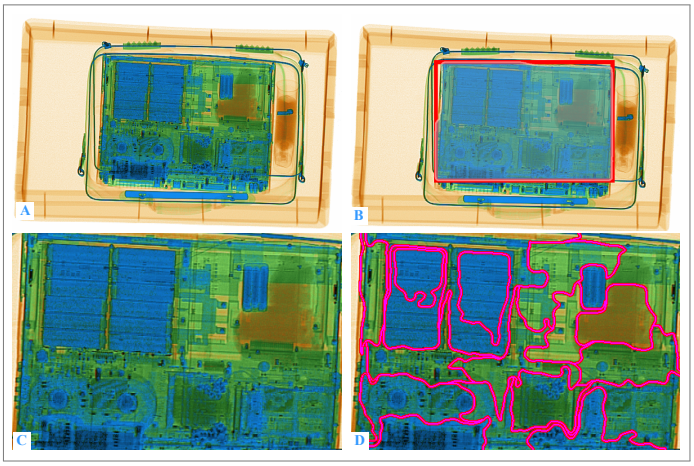

Fig. 1. Exemplar 2D X-ray imagery (A) used for object level anomaly detection (B/C) via mask R-CNN segmentation and sub-component level anomaly detection (D) via joint object over-segmentation and classification.

Earlier superpixel algorithms can be classified into clustering and graph based strategies. Clustering strategies [10], [11], leverage traditional clustering techniques such as k-means for superpixel segmentation. Graph based approaches [12], [13] define the superpixel over-segmentation as a graph partitioning problem, in which nodes are represented by pixels and the edges denote the strength of connectivity between adjacent pixels. Most of these methods rely on traditional hand-crafted features and do not use deep CNN techniques. More recent approach, such as in [14], deep features are used for superpixel segmentation bypassing the gradients through non-differentiable superpixel algorithms. CNN based unsupervised data clustering approaches are proposed [15], [16] in literature. A deep embedded clustering framework, for simultaneously learning feature representations and cluster assignments is proposed in the work of [16]. A more recently, an end-to-end deep learning-based clustering algorithm called Superpixel

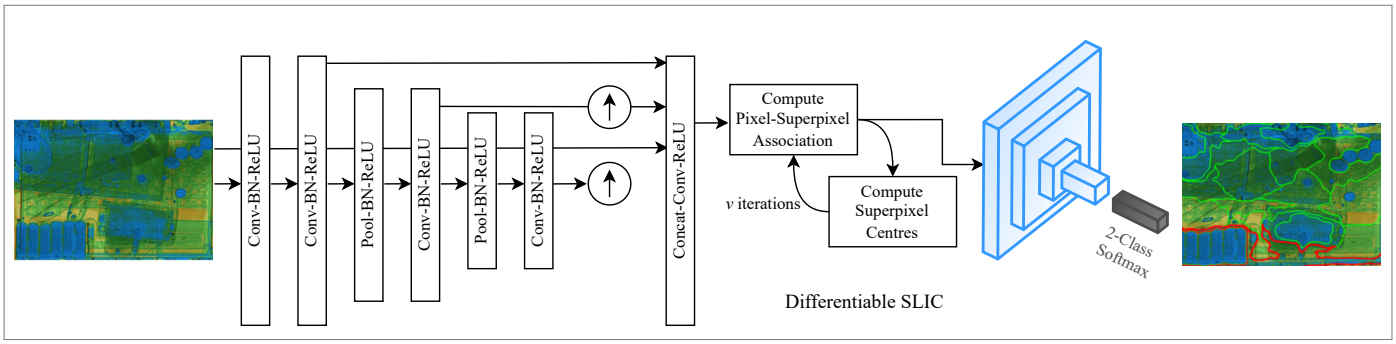

Fig. 2. End-to-end joint segmentation and classification CNN architecture. Each segment (green/red contours) is extracted using differentiable SLIC prior to classification.

Sampling Networks (SSN) [17] is developed for superpixel segmentation task where it can use image-specific constraints. However, none of the above discussed works is targeted for X-ray image application specific tasks. Following the work in [18], in our work we leverage the use of superpixels [10], [17] within X-ray security baggage image classification with prior object segmentation [19] as an enabler to CNN based end-to-end joint sub-component level anomaly detection within X-ray security imagery.

In this work, we evaluate two automatic segmentation strategies for intra-object anomaly detection in X-ray security imagery, as illustrated in the Figure 1A:-

- First, object level segmentation is performed (Figure 1B $\rightarrow$ Figure 1C).
- Secondly, end-to-end joint segmentation and classification CNN architecture is proposed (Figure 1C $\rightarrow$ Figure 1D) for anomaly detection as a binary, $\{anomaly, benign\}$, classification task.
- Additionally, we study the impact of dual-energy X-ray imagery (pseudo-colour, high, low and effective-Z) for anomaly detection task.

II. Proposed Approach

We outline the approach of this paper in the following section: object level localisation using Mask R-CNN [19] in Section II-A, followed by CNN based end-to-end segmentation and classification strategy in Section II-B.

A. Object Level Localisation

We consider contemporary CNN architectures, such as Mask R-CNN [19], Faster R-CNN [20], for object detection and segmentation task to explore their applicability for generalised object detection/instance segmentation tasks within the context of X-ray security baggage imagery. Mask R-CNN [19] relies on region proposals followed by ROI-Pooling to produce standard-sized outputs, which include pixel-wise image mask of a detected object, suitable for input into a secondary classifier. It addresses feature map misalignment of Faster R-CNN [20] by incorporating bilinear boundary interpolation. Mask R-CNN combines object localisation with instance segmentation of the object in the image (Figure 1B $\rightarrow$ Figure 1C). This architecture [19] is evaluated over an electronics and electrical items packed within cluttered X-ray security baggage, for anomaly detection task.

B. Joint Segmentation and Classification via CNN

For object over-segmentation (Figure 1D), we apply deep CNN based object sub-component level segmentation method, Super Sampling Network (SSN) [17] combined with classification network for $\{anomaly, benign\}$ classification. At the core of SSN is a differentiable clustering technique, which is inspired by Simple Linear Iterative Clustering (SLIC) [10] approach. SLIC performs iterative clustering, where initially image is segmented into roughly equal sized segments. To measure the similarity between the segments, it introduces a new distance metric which considers the size of the segment. It takes the user define number of approximately equally-sized superpixel K. For an image with N pixels, the approximate size of each superpixel will be N/K.

The core of part of SLIC [10] is iterative clustering, which is non-differentiable due to the computation of pixel to superpixel associations involving nearest neighbour operation. Instead of computing hard pixel-superpixel associations, SSN approach computes soft-associations between pixels and superpixels, which makes it differentiable.

Figure 2 depicts the joint segmentation and classification architecture. In SSN, the CNN for feature extraction is composed of a series of convolution layers interleaved with batch normalisation and ReLU activation. The max-pooling layer down samples the input by factor 2, after 2^{nd} and 4^{th} convolutional layer output. Convolution filter size of 3×3 is used with the number of output channels set to 64 in each layer. Other CNN architectures can also be integrated easily in this framework. The final features are passed onto the differentiable SLIC module, which iteratively updates the pixel-superpixel association and computes superpixel centres. We use reconstruction loss (l_{rcon}), which is cross-entropy loss, and compactness loss (l_{comp}) to reduce the spacial variance in each superpixel cluster. The total loss is the sum of the above two loss functions ($L = l_{rcon} + \lambda l_{comp}, where \ \lambda = 1e-4$).

Followed by this stage, the superpixels are fed into a binary classifier (as shown in the Figure 2) for anomaly detection in each segment. Each segmented image region, sub-component level segmentation, is subsequently classified

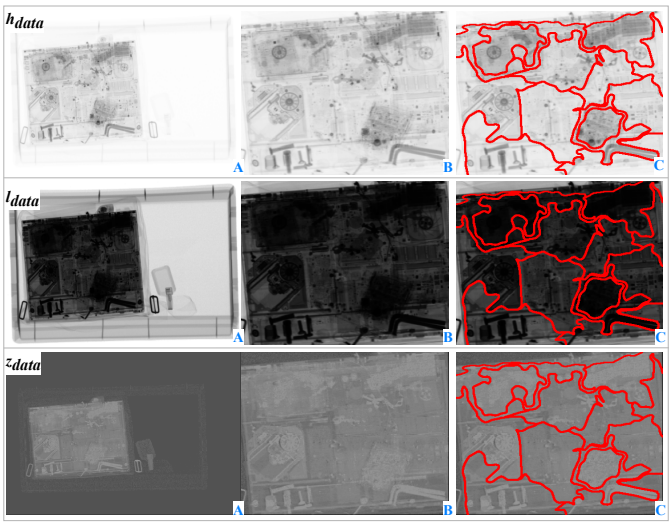

Fig. 3. Exemplar high, low, and effective-Z X-ray imagery from *DEEi* dataset (A) used for object level (B) and sub-component level segmentation (C) $\{anomaly, benign\}$ classification.

using a deep CNN architecture model formulated as a binary, $\{anomaly, benign\}$, classification task.

SqueezeNet [21] is a small network architecture that uses many 1-by-1 filters to aggressively reduce the number of weights. It offers equivalent accuracy to the AlexNet [5] yet operating with $50\times$ fewer parameters.

VGG [22] is a seminal network architecture that consists of several deep convolutional layers (e.g. $11-19$ weight layers), with a fixed kernel size (3×3) (convolution stride is set to 1), stacked on top of each other in increasing depth, which shows notable performance improvements on prior architecture [5], [23].

ResNet [6] solves the issue of vanishing gradient present in the forward feed and backward propagation processing in previous CNN architectures by introducing skip connection, parallel to the regular convolutional layers ($18-152$ in depth).

III. Evaluation

This section presents the dataset used, the implementation details, and the results of our experiments.

A. Experimental Setup

Our experimental setup comprises of following dataset.

DEEi. The dataset (Durham Electrical and Electronics Items) is constructed using a 2D X-ray scanner with associated pseudo-colour materials mapping via dual-energy. All X-ray imagery is gathered locally by using a dual-energy X-ray scanner (Gilardoni FEP ME 640 AMX) [24]. The dataset consists of large consumer electronics (e.g. laptop, mobile) and electrical (e.g. iron, hairdryer, toaster) items with and without anomaly (e.g. marzipan, screws, metal plates, sharps, etc.) concealment present as illustrated in Figures 1A. Our dataset consists of disparate models and shapes of electronics items, where the inserted anomaly concealment might be challenging to identify. In total, we use $7,022$ X-ray imagery ($70:30$ data split) for our experiment. Additionally, we access the dual-energy X-ray ('raw'), data (three types - high (h_{data}), low (l_{data}) energy, and effective-Z (z_{data}) response (Figure 3)) from the X-ray scanner [24] for $\{anomaly, benign\}$ classification.

Training for all architecture variants is performed via transfer learning using stochastic gradient descent with a momentum of 0.9, a learning rate of 0.0002, a batch size of 64. All networks are trained on NVIDIA 1080Ti GPU via the PyTorch framework [25].

B. Evaluation of Object and Sub-component Level Classification

Our model performances are evaluated in terms of Accuracy (A), Precision (P), F-score (F1), True Positive (TP%), and False Positive (FP%), as presented in the Tables I and II where we additionally compare our 'raw' X-ray data approach to the use of the pseudo-colour imagery conventionally used in automated object detection studies in X-ray images [3], [8], [26].

TABLE I
Object level segment and classification using varying CNN architectures with pseudo-colour and 'raw' X-ray data.

	Data	Network	A	P	F1	TP(%)	FP(%)
Object level segmentation	*pseudo-colour*	SqueezeNet	0.83	0.78	0.82	93.14	26.97
		VGG-16	0.76	0.69	0.75	94.26	39.47
		ResNet$_{50}$	**0.86**	**0.84**	**0.84**	**97.29**	**16.59**
	h_{data}	SqueezeNet	0.82	0.76	0.81	92.67	28.01
		VGG-16	0.75	0.67	0.73	96.76	45.46
		ResNet$_{50}$	0.85	0.80	0.85	94.83	22.24
	l_{data}	SqueezeNet	0.75	0.69	0.75	88.64	35.28
		VGG-16	0.71	0.63	0.68	97.18	55.01
		ResNet$_{50}$	0.81	0.80	0.64	84.84	20.76
	z_{data}	SqueezeNet	0.76	0.73	0.65	80.31	21.32
		VGG-16	0.68	0.66	0.64	84.58	47.73
		ResNet$_{50}$	0.84	0.80	0.81	90.62	20.45

In the object-level segmentation strategy (Table I), where the target object is first detected, localised and isolated via segmentation prior to binary classification, the maximum accuracy is achieved with ResNet$_{50}$ (A: 0.86, TP: 97.29% - Table I, *pseudo-colour*), due to the focused feature representation. We observe that the use of 'raw' X-ray data achieves good true positive (TP: 97.18% - Table I, l_{data}), but suffers with relatively high false positive rate. The lowest FP is 16.59% (with ResNet$_{50}$ - Table I, *pseudo-colour*), but fails to outperform object sub-component level strategy (FP: 4.54% - Table II, *pseudo-colour*).

From the results presented in Tables I and II, we observe from the two strategies considered that the joint sub-component level segmentation and classification strategy for anomaly detection via ResNet$_{50}$, offers significantly superior anomaly detection performance (A: 0.97, TP: 98.99%, FP: 4.54% - Table II, *pseudo-colour*) compared to an object level segmentation strategy overall (Table I). Although SqueezeNet

TABLE II
OBJECT SUB-COMPONENT LEVEL JOINT SEGMENTATION AND CLASSIFICATION WITH PSEUDO-COLOUR AND 'RAW' X-RAY DATA.

	Data	Network	A	P	F1	TP(%)	FP(%)
Sub-component level segmentation	*pseudo-colour*	SqueezeNet	0.95	0.92	0.94	99.10	8.90
		VGG-16	0.93	0.91	0.93	95.89	8.55
		ResNet$_{50}$	**0.97**	**0.95**	**0.97**	98.99	**4.54**
	h_{data}	SqueezeNet	0.96	0.94	0.96	98.86	6.12
		VGG-16	0.96	0.94	0.96	98.71	6.01
		ResNet$_{50}$	0.96	0.94	0.96	99.79	6.16
	l_{data}	SqueezeNet	0.93	0.91	0.93	96.85	9.53
		VGG-16	0.95	0.93	0.98	98.76	7.09
		ResNet$_{50}$	0.96	0.93	0.95	98.64	6.37
	z_{data}	SqueezeNet	0.90	0.87	0.89	95.28	15.06
		VGG-16	0.95	0.93	0.95	97.18	6.52
		ResNet$_{50}$	0.96	0.94	0.96	98.99	5.93
	hlz_{data}	SqueezeNet	0.96	0.94	0.96	99.74	6.26
		VGG-16	0.96	0.94	0.96	99.75	6.03
		ResNet$_{50}$	**0.97**	0.94	0.94	**100**	6.07

achieves the maximum true positive (TP: 99.10%), but it suffers relatively high false positive (FP: 8.90%) when using pseudo-colour X-ray imagery as presented in Table II - *pseudo-colour*. When we use 'raw' X-ray data for classification, we achieve the highest true positive (TP: 100%, A: 0.97, Table II, hlz_{data}) with ResNet$_{50}$ and combination of high, low and effective-Z energy data, but this is effected by the relatively high false positive rate (FP: 6.07%, Table II, hlz_{data}). Overall, ResNet$_{50}$ performs the best across all the three (high, low, effective-Z and combination) 'raw' X-xay data with TP $> 99\%$. Overall the sub-component level segmentation provides higher granularity information compared to the object level segmentation, henceforth achieves the best performance.

To our knowledge, the proposed work on joint end-to-end object sub-component level segmentation and classification is one of the first works for the anomaly detection task. Therefore, we consider two segmentation strategies, i.e., object-level (using Mask R-CNN)vs sub-component level to compare the results (Tables I, II).

Examples of the detection (object level segmentation via Mask R-CNN) and classification of the consumer electrical and electronics items containing an anomaly are depicted in Figure 4A. Figure 4B illustrates exemplary qualitative results of joint sub-component segmentation and classification of electrical and electronics items, where red colour indicates anomalous region while green represents benign sub-components. Examples of anomaly detection using 'raw' energy X-ray data are depicted in the Figure 5. The benefit of using sub-component level segmentation is it's capability to provide precise localisation of the anomalous region within complex object (Figures 4B, 5).

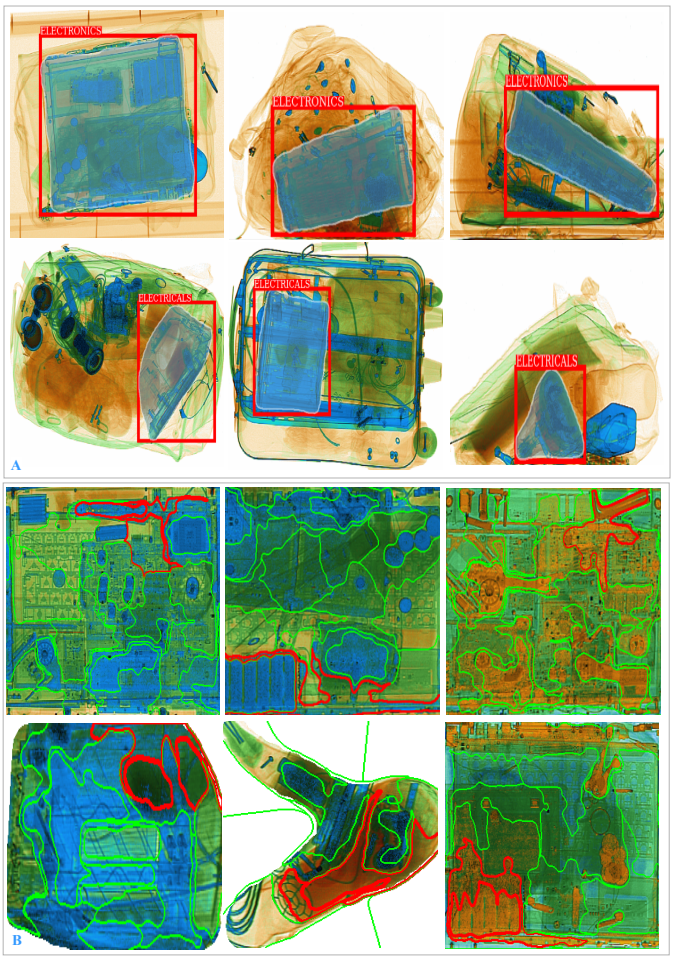

Fig. 4. Examples of $\{anomaly, benign\}$ detection and classification from *DEEi* dataset: (A) object level segmentation and (B) joint object sub-component segmentation (contours) and classification (green: benign, red: anomaly).

IV. CONCLUSION

We evaluate the performance impact of two different strategies, object segmentation, and object sub-component segmentation, for concealed threat/anomaly detection within consumer electrical/electronics item using deep CNN based end-to-end architecture. Our experimental results exhibit that the best performance ($> 99\%$ TP and $\sim 5\%$ FP) is achieved with object sub-component segmentation strategy on CNN classifier using 'raw' X-ray data (a combination of high, low energy and effective-Z). To the best of our knowledge, the use of 'raw' X-ray imagery for anomaly detection tasks is one of the first and novel of its kind. Therefore, our study possibly opens up a plethora of broad research interests in the X-ray imagery domain. Within the context of electrical and electronics items, this work offers the automatic first-stage screening of aviation baggage for anomalous item detection at the component level as an indicator of potential threat presence. In our future work on anomaly detection, we primarily focus on combining multiple data (a combination of pseudo-colour and high, low energy and effective-Z) to achieve higher accuracy and lower false positive. Additionally, we will target varied electronic and electrical items across a full range of operational X-ray characteristics.

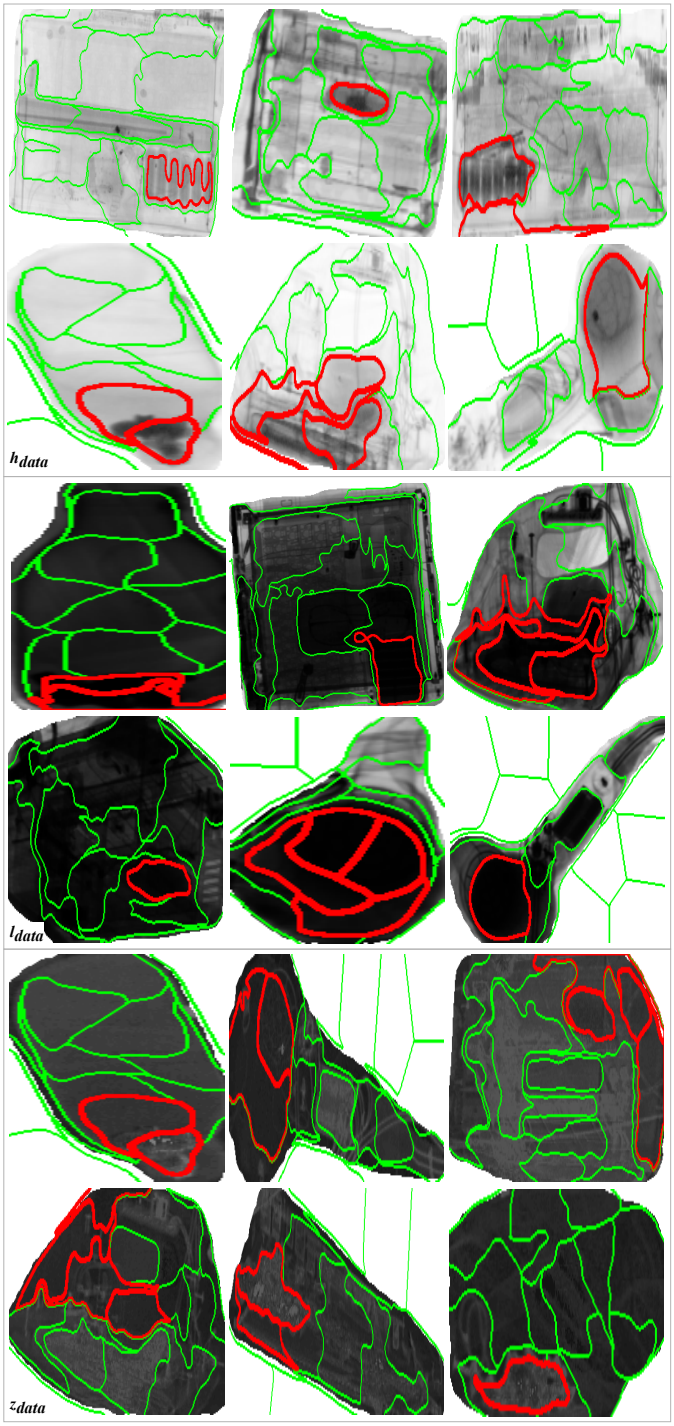

Fig. 5. Examples anomaly detection using high, low, and effective-Z X-ray imagery ('raw') from *DEEi* dataset: object sub-component segmentation (contours) and classification (green: benign, red: anomaly).

REFERENCES

[1] D. Turcsany, A. Mouton, and T. P. Breckon, "Improving feature-based object recognition for x-ray baggage security screening using primed visualwords," in *IEEE Int. Conf. on Industrial Technology*, Feb 2013, pp. 1140–1145.

[2] M. Bastan, W. Byeon, and T. M. Breuel, "Object recognition in multi-view dual energy x-ray images." in *Proc. British Machine Vision Conference*, vol. 1, no. 2, 2013, p. 11.

[3] S. Akcay and T. P. Breckon, "An evaluation of region based object detection strategies within x-ray baggage security imagery," in *Proc. Int. Conf. on Image Processing*, 2017, pp. 1337–1341.

[4] S. Akcay, M. Kundegorski, C. Willcocks, and T. Breckon, "On using deep convolutional neural network architectures for automated object detection and classification within x-ray baggage security imagery," *IEEE Transactions on Information Forensics Security*, vol. 13, no. 9, pp. 2203–2215, 2018.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. of the Int. Conf. on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[7] L. D. Griffin, M. Caldwell, J. T. A. Andrews, and H. Bohler, ""Unexpected item in the bagging area": Anomaly detection in x-ray security images," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 6, pp. 1539–1553, 2019.

[8] S. Akcay, A. A. Abarghouei, and T. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in *Proc. Asian Conf. on Computer Vision*. Springer International Publishing, 2018.

[9] S. Akcay, A. Atapour-Abarghouei, and T. Breckon, "Skip-ganomaly: Skip connected and adversarially trained encoder-decoder anomaly detection," in *Proc. Int. Joint Conf. on Neural Networks*. IEEE, 2019, pp. 1–8.

[10] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2274–2282, 2012.

[11] Z. Li and J. Chen, "Superpixel segmentation using linear spectral clustering," in *Proc. Conf. on Computer Vision and Pattern Recognition*, 2015, pp. 1356–1363.

[12] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proc. of Ninth Int. Conf. on Computer Vision*, 2003, p. 10.

[13] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. journal of computer vision*, vol. 59, no. 2, pp. 167–181, 2004.

[14] W.-C. Tu, M.-Y. Liu, V. Jampani, D. Sun, S.-Y. Chien, M.-H. Yang, and J. Kautz, "Learning superpixels with segmentation-aware affinity loss," in *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2018, pp. 568–576.

[15] A. Rasmus, M. Berglund, M. Honkala, H. Valpola, and T. Raiko, "Semi-supervised learning with ladder networks," in *Advances in neural information processing systems*, 2015, pp. 3546–3554.

[16] J. Xie, R. Girshick, and A. Farhadi, "Unsupervised deep embedding for clustering analysis," in *Proc. Int. Conf. on machine learning*, 2016, pp. 478–487.

[17] V. Jampani, D. Sun, M.-Y. Liu, M.-H. Yang, and J. Kautz, "Superpixel sampling networks," in *Proc. of the European Conf. on Computer Vision*, 2018, pp. 352–368.

[18] J. Zhang, L. Zhang, Z. Zhao, Y. Liu, J. Gu, Q. Li, and D. Zhang, "Joint shape and texture based x-ray cargo image classification," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2014, pp. 266–273.

[19] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. of the IEEE Int. Conf. on Computer Vision*, 2017, pp. 2961–2969.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[21] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5 mb model size," *preprint arXiv:1602.07360*, 2016.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. on Learning Representations*, 2015.

[23] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Computer Vision – ECCV 2014*, 2014, pp. 818–833.

[24] "Gilardoni," www.gilardoni.it/en/security, accessed: 2022-07-06.

[25] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," 2017.

[26] Y. A. Gaus, N. Bhowmik, S. Akçay, P. Guillen-Garcia, J. Barker, and T. Breckon, "Evaluation of a dual convolutional neural network architecture for object-wise anomaly detection in cluttered x-ray security imagery," in *Proc. Int. Joint Conf. on Neural Networks*, 2019.