On the Impact of Object and Sub-component Level Segmentation Strategies for Supervised Anomaly Detection within X-ray Security Imagery

Neelanjan Bhowmik[1], Yona Falinie A. Gaus[1], Samet Akçay[1], Jack W. Barker[1], Toby P. Breckon[1,2]
Department of {Computer Science[1] | Engineering[2]}, Durham University, UK

***Abstract*—X-ray security screening is in widespread use to maintain transportation security against a wide range of potential threat profiles. Of particular interest is the recent focus on the use of automated screening approaches, including the potential anomaly detection as a methodology for concealment detection within complex electronic items. Here we address this problem considering varying segmentation strategies to enable the use of both object level and sub-component level anomaly detection via the use of secondary convolutional neural network (CNN) architectures. Relative performance is evaluated over an extensive dataset of exemplar cluttered X-ray imagery, with a focus on consumer electronics items. We find that sub-component level segmentation produces marginally superior performance in the secondary anomaly detection via classification stage, with true positive of $\sim 98\%$ of anomalies, with a $\sim 3\%$ false positive.**

***Index Terms*—X-ray imagery, electronics item, superpixel, anomaly detection, CNN, classification**

I. Introduction

X-ray baggage security screening is widely used to maintain aviation and transport security, itself posing a significant image-based screening task for human operators reviewing compact, cluttered and highly varying baggage contents within limited time-scales. With both increased passenger throughput in the global travel network and an increasing focus on wider aspects of extended border security (e.g. freight, shipping, postal), this poses both a challenging and timely automated image classification task.

Prior work in the field has notably concentrated on the shaped-based detection of both threat and contraband (undeclared) items within X-ray imagery achieving both high detection performance with low false positive reporting [1]–[4]. However, such approaches are insufficient when dealing with the detection of unknown anomalous items or materials potentially concealed within complex items such as consumer electronic devices.

Whilst existing security scanners use dual-energy X-ray for materials discrimination, and highlight specific image regions matching existing threat material profiles [5], [6], the detection of generalized anomalies within complex items remains challenging [7] (e.g. Figure 1).

Within machine learning, anomaly detection involves learning a pattern or distribution of normality for a given data source and thus detecting significant deviations from this norm [8]. Anomaly detection is an area of significant interest within computer vision, spanning biomedical imaging [9] to video surveillance [10]. In our consideration of X-ray security imagery, we are looking for abnormalities that indicate concealment or subterfuge whilst working against a real-world adversary who may evolve their strategy to avoid detection. Such anomalies may present (or conceal) themselves within appearance space in the form of an unusual shape, texture or material density (i.e. dual energy X-ray colour) [11]. Alternatively they may present themselves in a semantic form, where the appearance of unfamiliar objects either globally or locally within the X-ray image [12].

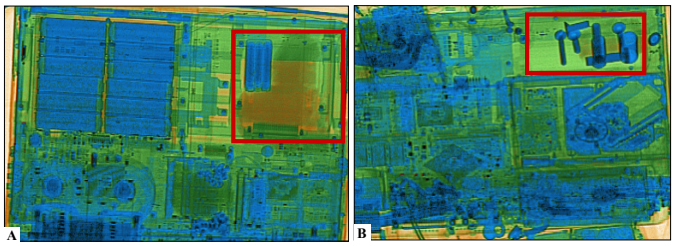

Fig. 1. Exemplar consumer electronics item within X-ray security imagery with both material (red box, A) and embedded object (red box, B) anomaly present.

Prior work on appearance and semantic anomaly detection, has considered unique feature representation as a critical component for detection within cluttered X-ray imagery [13]. Early work on anomaly detection in X-ray security imagery [1], implements block-wise correlation analysis between two temporally aligned scanned X-ray images. More recently [14], anomalous X-ray items within freight containers have been detected using auto-encoder networks, and additionally via the use convolutional neural network (CNN) extracted features as a learned representation of normality across stream-of-commerce parcel X-ray images [13]. In a similar vein, the work of [15] focuses on the use of a novel adversarial training architecture to detect anomalies as high reconstruction errors produced from a generator network adversarially trained on non-anomalous (benign) stream-of-commerce X-ray imagery only.

However, the majority of this prior anomaly detection work is focused at the image or object level, where anomaly presence is clear in appearance or semantic space, by asking the global question - *is the image anomalous?*

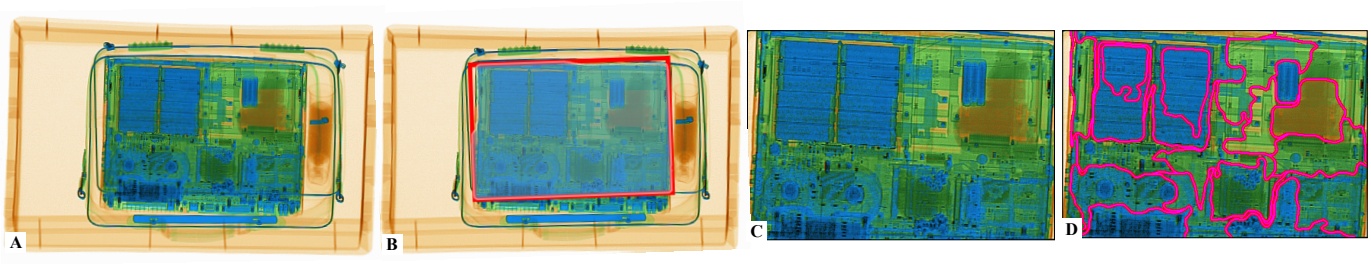

Fig. 2. Exemplar X-ray imagery (A) used for object level anomaly detection (B/C) via mask R-CNN segmentation and sub-component level anomaly detection (D) via superpixel over-segmentation.

These approaches [13]–[15], fail to address the fact that anomaly presence maybe subtle and concealed (i.e. present) within a semantically benign object itself (e.g. Figure 1 A/B). In this case, we wish to ask a highly localised question - *is this part of this complex object within the image anomalous ?*

In order to address this issue, we consider the task of image segmentation - if we first segment a class of object from the image, then potentially segment that object into its sub-components how well can this issue of subtle and concealed anomaly detection be addressed.

To these ends, we introduce a side-by-side comparison of both object and sub-component level segmentation strategies for this case of intra-object anomaly detection. While anomaly detection at an object level is more common, the detection in sub-component level is still at infancy. The key concept is that whilst subtle localised anomalies maybe difficult to detect via an image level anomaly detection approach, we can instead target object level or sub-component level anomaly detection in isolation. Hence a more general learning-driven approach can developed at the object or sub-component level instead of tackling global signatures across all possible objects - *and thus being able to tell if they are anomalous or benign in appearance / semantic space*.

Following the work in Zhang et al, [16] where they leverage the use of superpixels [17] within X-ray cargo image classification, we complement such approach with prior object segmentation [18] as an enabler to sub-component level anomaly detection within X-ray security imagery. Using contemporary object segmentation via mask Region-based CNN (R-CNN) [18] and Simple Linear Iterative Clustering (SLIC) [17] superpixels, we evaluate alternate strategies for the detection of subtle intra-object anomalies at either a generalised object level or sub-component level segmentation strategy, thus facilitating effective anomaly detection independent of resolute object classification (Section II). Our work is evaluated over a range of large consumer electronics items with and without intra-object anomaly presence (Section III).

II. Proposed Approach

Our approach considers two automatic segmentation strategies for intra-object anomaly detection in X-ray security imagery (Sec. II-A), as illustrated in the Figure 2A:- first, object level segmentation is performed (Figures 2B $\rightarrow$ 2C) and secondly, sub-component level segmentation is performed (Figure 2C $\rightarrow$ 2D). This is followed by secondary scale-specific variants to contemporary deep CNN architectures for final anomaly detection as a binary, $\{anomaly, benign\}$, classification task (Sec. II-B).

A. Segmentation Strategies

Object Level Segmentation: Our first segmentation strategy builds upon the Faster R-CNN [19] X-ray security image specific work of [4], to augment this model by adding two additional convolutional layers to construct a object boundary segmentation mask, following the Mask R-CNN concept of [18]. This is performed by adding an additional branch to Faster R-CNN that outputs an additional image mask indicating pixel membership of a given detected object. Mask R-CNN [18] also addresses feature map misalignment, found in Faster R-CNN [19] for higher resolution feature map boundaries, via bilinear boundary interpolation. Our Mask R-CNN is applied to an input X-ray image (Figure 2A) with segmented object (Figure 2B then isolated from the image for subsequent object level anomaly detection (Figure 2C).

Sub-component Level Segmentation: Our second segmentation strategy uses image over-segmentation via Simple Linear Iterative Clustering (SLIC) [17] superpixels. It performs iterative clustering in a similar manner to k-means, where the image is segmented into approximately equally-sized superpixels, whose total number k is user-defined. SLIC represents each pixel in $\mathbb{R}^5$, defined by the $\{L, a, b\}$ values of CIELAB colour space and the (x, y) pixel coordinate. Instead of using Euclidean distance, SLIC introduce a new distance measure that considers superpixel size. SLIC takes as input a desired number of approximately equally-sized superpixel K, and for the images with N pixels, with the approximate size of each superpixel will be N/K. Each of every approximately equally-sized superpixels, there will be a superpixel center at every grid interval $S = \sqrt{N/K}$. Let $\left[l_i, a_i, b_i, x_i, y_i\right]^T$ be the five dimensional point of a pixel, cluster center C_k should be in the same form as $\left[l_k, a_k, b_k, x_k, y_k\right]^T$. The distance measure D_k is defined as:

$$\begin{aligned} d_{lab} &= \sqrt{(l_k - l_i)^2 + (a_k - a_i)^2 + (n_k - b_i)^2} \\ d_{xy} &= \sqrt{(x_k - x_i)^2 + (y_k - y_i)^2} \\ D_s &= d_{lab} + \frac{m}{S} d_{xy} \end{aligned} \tag{1}$$

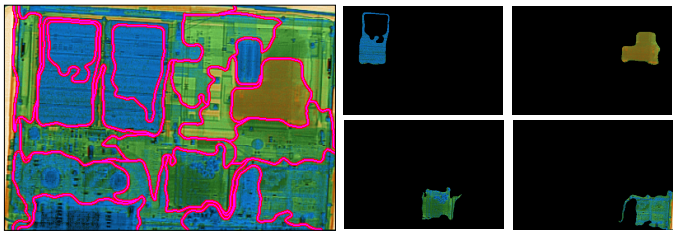

Fig. 3. Sub-component level object segmentation: each segment (pink contours) is extracted prior to CNN classification.

where D_s is the sum of the lab distance and the xy plane distance $normalized$ by the grid interval S. Variable m is introduced to control the compactness of the superpixel with the local convexity or concavity shape of each superpixel dependant on *m* (low *m* reduces the influence of coordinate information while for a high *m* each superpixel will approximate a square shape). Our $m = 20$ choice (by taking consideration of the size of the object present in an image), results in a set superpixel region conforming to convex and concave image shape boundaries as illustrated in the Figure 3.

B. Secondary Classification

Each segmented image region, from object level or sub-component level segmentation, is subsequently classified using a deep CNN architecture model formulated as a binary, $\{anomaly, benign\}$, classification task. Three contemporary generalised CNN architectures plus leading fine-grain CNN classification approaches [20]–[22], specifically targeting the sub-categorization of pre-determined object types, are considered to form the basis of our anomaly detection study.

VGG-16 [23] is a seminal network architecture that consists of 16 deep convolutional layers, with a fixed kernel size of 3, stacked on top of each other in increasing depth.

SqueezeNet [24] 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 [25] yet operating with $50\times$ fewer parameters.

ResNet-50 [26] 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 numbering 50 in depth.

Fine-grain Classification [20]–[22] is put into effect as we can consider the task of anomaly detection in our case as a fine-grained image classification (FGIC) problem. The Xray screening imagery used, has very subtle differentiating factors in the sub-component of a given object (e.g laptop, bottle) and as such a fine grained approach should be used to detect finer class-specific discriminatory patches within objects [27], [28]. *Bilinear Convolutional Neural Network (BCNN)* [20] utilises a dual VGG-16 architecture in parallel with each stream implements uncommon, trivial elements of convolution and max pooling thus allowing focus on two separate distinct parts of the object. The two streams are concatenated into a bilinear vector using sum pooling over the outputs of both streams. This is then used in the final classification by feeding into the linear layers of the network, and finally a softmax layer to gain a probabilistic output of the most likely classifications for the image.

Multi-Attention (MA) [21] optimises part attentions of four distinct regions of an image using the feature channels in a VGG-16 architecture. This allows the network to focus on discriminative factors present in object parts, and use this in the final classification. Each of the four layers produces a classification at the end of the network linear layers which is then grouped by channel grouping loss in order to generate a final classification for a given object.

Discriminative Filter Bank (DFL) approach [22], heightens the mid-level network in a VGG-16 based architecture, by learning a collection of 1×1 convolution filters known as a filter bank (FB), and a 92×92 with stride 8 to preserve global shape and appearance dependency in the image data [22]. These filters when properly initialised and successfully learned can respond to discriminative regions when convoluted over the image.

When applied to the challenge of X-ray security screening, for the binary classification problem $\{anomaly, benign\}$, the models should be able to recognise much subtler visual differences and locations of such parts within object sub-components which will ultimately lead to more reliable classification.

Each CNN architecture is trained via a transfer learning approach, with pre-training on the 1000-class ImageNet [29] object classification problem, for our final two-class (binary) X-ray imagery classification problem, $\{anomaly, benign\}$. For both object level and sub-component level segmentation our resulting image segments are padded and re-scaled to a common reference dimension (objects: 224×224; sub-components (superpixels): 190×150). Dataset imbalance, a common problem for anomaly detection problems where anomalous examples can be scarce and challenging to obtain, is addressed by up-sampling the anomalous class with the lesser volume of samples. In total training is performed over a dataset of 14,964 X-ray imagery ($70 : 30$ data split) and testing reported over a dataset of 7,878 X-ray imagery (50%: anomalous and 50%: benign) containing consumer electronics items.

Training is performed via transfer learning using stochastic gradient descent with a momentum of 0.9, a learning rate of 0.001, a batch size of 64 and categorical cross-entropy loss. All networks are trained on NVIDIA 1080 Ti GPU via PyTorch [30].

III. Evaluation

Our evaluation considers the comparative performance of: (a) object level segmentation followed by anomaly detection via CNN classification (i.e., anomaly present in object as a whole - $\{anomaly, benign\}$) and (b) sub-component level segmentation followed by anomaly detection via CNN classification (i.e., anomaly present in image sub-component patches, i.e., superpixels - $\{anomaly, benign\}$). We consider statistical

TABLE I
SUB-COMPONENT LEVEL SEGMENT CLASSIFICATION OF SECURITY X-RAY IMAGERY USING VARYING CNN ARCHITECTURES.

Strategy	Model	Architecture	A	P	F1	TP(%)	FP(%)
Sub-component level segmentation	Binary Classification via CNN	ResNet-18 [26]	97.10	95.40	97.00	98.89	4.69
		ResNet-50 [26]	97.20	95.50	97.10	98.99	4.54
		SqueezeNet [24]	95.10	92.60	94.70	**99.10**	8.90
		VGG-16 [23]	93.70	91.80	93.30	95.89	8.55
	Fine-Grain Classification	BCNN [20]	97.54	95.53	97.49	95.49	4.30
		MA [21]	97.68	95.81	97.63	96.32	4.19
		DFL [22]	**97.91**	**96.40**	**97.87**	98.20	**3.50**

TABLE II
OBJECT LEVEL SEGMENT CLASSIFICATION OF SECURITY X-RAY IMAGERY USING VARYING CNN ARCHITECTURES.

Strategy	Model	Architecture	A	P	F1	TP(%)	FP(%)
Object level segmentation	Binary Classification via CNN	ResNet-18 [26]	86.20	80.60	76.90	95.42	21.13
		ResNet-50 [26]	86.20	**84.50**	84.50	**97.29**	16.59
		SqueezeNet [24]	83.40	78.10	82.20	93.14	26.97
		VGG-16 [23]	76.80	69.60	75.20	94.26	39.47
	Fine-Grain Classification	DFL [22]	**89.77**	83.70	**89.33**	83.70	**3.88**

Accuracy *(A)*, Precision *(P)*, F-score *(F1)*, True Positive *(TP)* and False Positive *(FP)* as presented in Tables I and II.

The X-ray security imagery dataset used for evaluation is obtained using a conventional 2D X-ray scanner with associated false colour materials mapping from dual-energy X-ray materials information [31]. It comprises large consumer electronics items (e.g., laptops) with and without intra-object anomaly concealment present. Anomaly concealments consist of marzipan, metal screws, metal plates, knife blades and similar inside the electronic items as illustrated in the examples of Figure 1 and Figure 2A/B.

Performance evaluation of the object level segmentation and component level segmentation approaches are performed over a set of $7,878$ images annotated with ground truth anomaly location gathered using local access to a dual-view X-ray cabin baggage security scanner.

From the results presented in Tables I and II, we can observe that a sub-component level segmentation strategy, supported by the secondary fine-grain CNN classification of DFL model [22], offers significantly superior anomaly detection performance (A: 97.91, TP: 98.20, FP: 3.50 - Table I) than an object level segmentation strategy overall (Table II). Furthermore, fine-grain CNN classification similarly offers the highest overall accuracy and lowest false positive rate (A: 89.77, FP: 3.88 - Table II) for object level segmentation. By contrast, the use of binary classification via a CNN offers superior performance for object level segmentation (Table II) in terms of higher accuracy supported primarily by higher true positive detection at the expense of false positive reporting. Second stage binary classification via CNN performed less well overall with the sub-component segmentation strategy (lower accuracy (A) caused by significantly higher false positive (FP) - Table II).

Fine grain classification model (DFL [22]) offer the lowest false positive and maximal accuracy for both segmentation strategies (Table I). We can deduce that increased levels isolation via segmentation to the sub-component level improves the performance of the discriminative feature space learnt by the fine-grain technique [20]–[22] whilst more classical object classification CNN architectures perform only marginally better on objects than sub-components (Table II).

Figure 4 illustrates the attention (red/pink patch with the highest focus) of the fine-grain DFL model [22] whilst trained on object-level and sub-component level segmentation data respectively. They are generated on the Rectified Linear Unit (ReLU) activations of the final layer before the fully connected layer in the VGG-16 [23] architecture of DFL. It is evident when inspecting these that the sub-level components (Figure 4 second column) show attention over the anomalous parts of the laptops in each image while the object-level analysis (Figure 4 first column) shows relatively sporadic sparse attention over the images. This provides qualitative visualization supporting the performance of the sub-component level segmentation strategy outperforming object level segmentation.

By enhancing the intermediate layer within the VGG-16 via a filter bank [22], we can hypothesise that this allows it to learn edge, corner and texture detail on specific sub-components at finer-level of the candidate region presented. As a result, the learned feature representation of anomaly against benign is highly discriminative leading to a significantly lower false positive than any other technique - at both the object and sub-component level (Tables I and II).

Binary classification via CNN using same VGG-16 architecture can achieve high true positive detection at the object level but at the expense of the highest false positive (TP: 94.25, FP: 39.47 - Table II). This can also be observed for the ResNet and

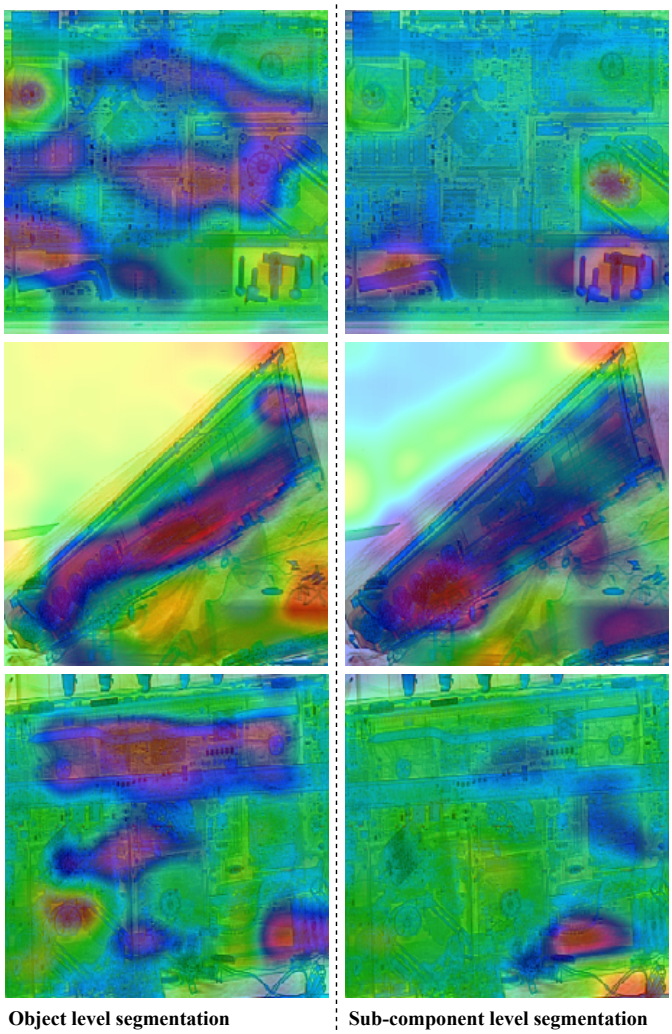

Fig. 4. Heatmap generated by convolutional activation map with DFL model [22] trained on object level segmentation images in first column and sub-component level segmentation images in second column. This shows where the model is looking (red/pink colour patch with primary focus regions) when selecting discriminative regions within the images using.

SqueezeNet architectures. For example, ResNet-50 achieves true positive of 97.29%, however suffering from high FP of 16.59% for object level segment classification (Table II).

Overall we observe that a sub-component level segmentation strategy, enabled via object segmentation via Mask R-CNN [18] and subsequent superpixel over-segmentation via SLIC [17], consistently outperforms an object level segmentation strategy (via Mask R-CNN [18] alone) when secondary region classification is performed using a specific fine-grain CNN variant [22]. The mean runtime for end-to-end $\{anomaly, benign\}$ classification strategy (object segmentation, followed by sub-component level segmentation and fine-grain classification) is ~ 500 milliseconds, which is within the belt speed (0.2meter/second) of standard X-ray scanner [32].

We primarily focus on supervised anomaly detection strategies and compared the performances amongst. Hence we do not include unsupervised or semi-supervised anomaly detection strategy [15] in our experiments and we believe it is not an equitable comparison between supervised and semi-supervised approaches. To the best our knowledge, the proposed work on $\{anomaly, benign\}$ classification within large consumer electronics items, using sub-component level segmentation strategy, is first of its kind. As there is no prior related work is available on the literature of X-ray security imagery (e.g. sub-component level segmentation classification), we are unable to compare our strategies with any existing algorithm and present our results as the benchmark.

Figure 5 shows exemplar qualitative results of sub-component level segmentation with per superpixel classification using the fine-grained DFL [22] approach where we can see the colour coded set of anomalous (red) as well as benign (green) sub-component regions within the pre-isolated object-level image region.

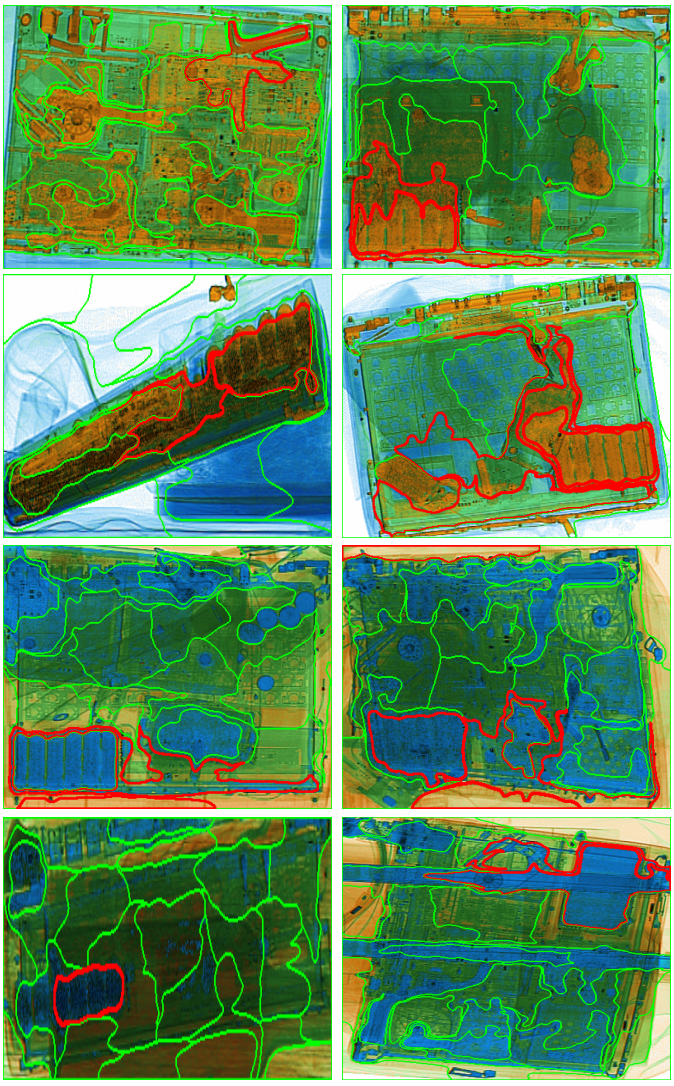

Fig. 5. Sub-component level segmentation via the use of SLIC approach [17] and classification via fine-grain DFL [22] applied in X- ray security imagery (red contour: anomaly, green contour: benign).

IV. Conclusion

We assess the performance impact of varying segmentation strategies, such as object level and sub-component level segmentation, for intra-object anomaly detection within the context of X-ray security imagery. Our experimental comparison demonstrates the superiority of a sub-component level segmentation approach in combination with a specific fine-grain CNN architecture achieving a performance accuracy of 97.91% with a notable 3.50% false positive rate for realistic anomaly concealment within representative consumer electronic items.

Future work will consider the conglomerate use of the multiple sub-component anomaly detection results in the robust determination of image-level anomaly vs. benign decision making for a broader range of object types.

Acknowledgements: Funding support - UK Department of Transport, Future Aviation Security Solutions (FASS) programme, (2018/2019).

References

[1] A. E. Y. Zheng, "A vehicle threat detection system using correlation analysis and synthesized x-ray images," in *Proc.SPIE*, vol. 8709, 2013, pp. 8709 – 8709 – 10.

[2] N. Jaccard, T. W. Rogers, and L. D. Griffin, "Automated detection of cars in transmission x-ray images of freight containers," in *2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug 2014, pp. 387–392.

[3] S. Akcay and T. P. Breckon, "An evaluation of region based object detection strategies within x-ray baggage security imagery," in *2017 IEEE International Conference on Image Processing (ICIP)*, Sep. 2017, pp. 1337–1341.

[4] 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*, vol. 13, no. 9, pp. 2203–2215, Sep. 2018.

[5] S. Singh and M. Singh, "Explosives detection systems (eds) for aviation security," *Signal Processing*, vol. 83, no. 1, pp. 31–55, 2003.

[6] K. Wells and D. Bradley, "A review of x-ray explosives detection techniques for checked baggage," *Applied Radiation and Isotopes*, vol. 70, no. 8, pp. 1729–1746, 2012.

[7] J. Skorupski and P. Uchroński, "Evaluation of the effectiveness of an airport passenger and baggage security screening system," *Journal of Air Transport Management*, vol. 66, pp. 53–64, jan 2018.

[8] A. Patcha and J. Park, "An overview of anomaly detection techniques: Existing solutions and latest technological trends," *Computer Networks*, vol. 51, no. 12, pp. 3448 – 3470, 2007.

[9] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Information Processing in Medical Imaging*. Cham: Springer International Publishing, 2017, pp. 146–157.

[10] B. R. Kiran, D. M. Thomas, and R. Parakkal, "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos," *Journal of Imaging*, vol. 4, no. 2, 2018.

[11] L. Greenemeier, "Exposing the weakest link: As airline passenger security tightens, bombers target cargo holds," Nov 2010. [Online]. Available: https://www.scientificamerican.com/article/aircraft-cargo-bomb-security/

[12] M. Brown, "Brothers plead not guilty to meat grinder bomb plot," May 2018. [Online]. Available: https://www.abc.net.au/news/2018-05-04/brothers-accused-of-plotting-to-blow-up-plane-plead-not-guilty/9726952

[13] L. D. Griffin, M. Caldwell, J. T. A. Andrews, and H. Bohler, "unexpected item in the bagging area," *IEEE Transactions on Information Forensics and Security*, pp. 1–1, 2018.

[14] J. T. A. Andrews, E. J. Morton, and L. D. Griffin, "Detecting Anomalous Data Using Auto-Encoders," *International Journal of Machine Learning and Computing*, vol. 6, no. 1, pp. 21–26, 2016.

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

[16] 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 *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*, June 2014, pp. 266–273.

[17] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Ssstrunk, "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, Nov 2012.

[18] 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.

[19] 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.

[20] T.-Y. Lin and S. Maji, "Improved bilinear pooling with cnns," in *Proceedings of the British Machine Vision Conference (BMVC)*, G. B. Tae-Kyun Kim, Stefanos Zafeiriou and K. Mikolajczyk, Eds. BMVA Press, September 2017, pp. 117.1–117.12.

[21] H. Zheng, J. Fu, T. Mei, and J. Luo, "Learning Multi-Attention Convolutional Neural Network for Fine-Grained Image Recognition," in *ICCV*, 2017.

[22] Y. Wang, V. I. Morariu, and L. S. Davis, "Learning a discriminative filter bank within a cnn for fine-grained recognition," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4148–4157, 2018.

[23] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *International Conference on Learning Representations*, 2015.

[24] F. N. Iandola, M. W. Moskewicz, K. Ashraf, S. Han, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <1mb model size," *CoRR*, vol. abs/1602.07360, 2016. [Online]. Available: http://arxiv.org/abs/1602.07360

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105.

[26] 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.

[27] G. V. Horn, S. Branson, R. Farrell, S. Haber, J. Barry, P. Ipeirotis, P. Perona, and S. Belongie, "Building a bird recognition app and large scale dataset with citizen scientists: The fine print in fine-grained dataset collection," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 595–604.

[28] J. D. Wegner, S. Branson, D. H. Hall, K. Schindler, and P. Perona, "Cataloging public objects using aerial and street-level images urban trees," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6014–6023, 2016.

[29] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Proc. Computer Vision and Pattern Recognition*, 2009, pp. 248–255.

[30] 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.

[31] A. Mouton and T. Breckon, "A review of automated image understanding within 3d baggage computed tomography security screening," *Journal of X-ray science and technology*, vol. 23, no. 5, pp. 531–555, 2015.

[32] "FEP ME 640 AMX," https://www.gilardoni.it/en/security/x-ray-solutions/automatic-detection-of-explosives/fep-me-640-amx/, accessed: 2019-10-14.