 DOI: 10.3217/978-3-85125-869-1-12

Benign Object Detection and Distractor Removal in 2D Baggage Scans

Anna Sebernegg[1] and Walter G. Kropatsch[2]

***Abstract*— Baggage screening contributes to security by helping to identify threats. However, the complexity of X-ray scans and the high intra-class variability make universal appearance-based threat detection difficult. Consequently, baggage inspection still relies on human operators, and further developments to assist them in their visual search tasks are desirable. This work proposes utilizing object detection as a diagnostic aid, where distractive benign objects are automatically detected and removed from the images through inpainting. The applied distractor removal successfully reduces visual saliency in benign regions and decreases the overall clutter of the scans.**

I. INTRODUCTION

Baggage inspection is increasingly automated, especially liquid and explosive detection systems have emerged [19]. Nevertheless, automatic appearance-based threat detection is hardly available due to the challenging nature of 2D baggage scans [1], which include high levels of clutter and overlapping objects due to tightly packed luggage [21], in- and out-of-plane rotations [5], and schemes to conceal prohibited items [11]. Therefore, human operators are still required to detect threats over visual search [10]. This task demands sustained attention over extended periods [19] and is negatively affected by several factors, such as the stressful environment [14] and complexity of baggage scans [18].

One possible way to support screeners and improve the visual search task could be to enhance the baggage scans by utilizing automatic object detection as a diagnostic aid. Like computer-aided detection systems used in the medical field [9], detected regions could be processed to focus the viewer's attention on critical content that requires further investigation. One potential application proposed in this work is to reduce the number of benign items that negatively contribute to the visual clutter by detecting and inpainting them automatically.

II. RELATED WORK

Extensive research is being conducted in appearance-based object detection within baggage scans [8]. However, the focus is on threats rather than benign items. *Image processing* is another broad field utilized in scans to improve readability [16], e.g., using material filters such as the organic-only filter mentioned by Michel et al. [16]. Saliency-driven image manipulation techniques such as distractor removal [6] or attention retargeting [12] are especially explored for photography and are less common in baggage screening.

*This work was not supported by any organization

[1,2]Pattern Recognition and Image Processing group, Technische Universität Wien, Favoritenstrasse 9-11, Vienna, Austria

[1]Email: e1526184@student.tuwien.ac.at

[2]Email: krw@prip.tuwien.ac.at

III. METHODOLOGY

This paper presents and experimentally evaluates a concept for automatic detection and removal of benign items from baggage scans. The goal is to reduce distractors to diminish visual clutter and shift saliency to other image regions. Therefore, object detection and distractor removal through inpainting is applied, as visualized in Fig. 1. The inpainting method should meet the following requirements:

- Reduce the overall visual clutter of the image
- Decrease salience in the inpainted benign regions
- Maintain or even increase salience in the rest of the image, especially in regions containing threats

Before applying inpainting, the regions of interest must be identified, which is done automatically by using a Convolutional Neural Network (CNN) for object detection that provides bounding boxes of the detected threats and benign items to a subsequent semantic segmentation performed in MATLAB. The object detector is received by applying transfer learning to the pre-trained EfficientDet model (D1) provided by the TensorFlow Object Detection API [3]. The database used for training and evaluation of the model consists of 3721 X-ray scans obtained from the public mono-energy X-ray database *GDXray* [15] and baggage scans created in cooperation with the CT Research Group at Wels Campus in upper Austria. The database is divided into a training set with 2938 images, a validation set with 632 images, and a test set with 151 images, whereby no images of single objects are included for testing. The final model can detect four threats and eight benign objects. Segmentation is performed by binarizing the image using MATLAB's implementation of Otsu's method [17] and morphological operations.

The following inpainting approaches are tested to find a suitable method for distractor removal, where the last three are provided functions by MATLAB [13]: *Uniform Inpainting* (inpainting with a uniform color from the background of the image), *Inpaint Coherent* (coherence transport based inpainting as described by Bornemann and März [2]), *Inpaint Exampler* (exemplar-based inpainting as described by Criminisi et al. [4]), and *Regionfill* (inpainting by inward interpolation from the outer pixels of the region [13]).

The effects of distractor removal on human visual attention are evaluated in two ways. Firstly, quadtree complexity as proposed by Jégou and Deblonde [20] is used to obtain the enhanced baggage scan's total visual clutter and compare it to the original image to determine if the distractor removal successfully reduces clutter. This method performs quadtree decomposition, where the number of cells in the resulting quadtree determines the global clutter value. Secondly, local saliency changes are measured using the Itti-Koch-Niebur Saliency Model (IKN) [7] provided by the *Saliency Model Implementation Library for Experimental Research* (SMILER) [22]. A saliency map, as shown in Fig. 2, is calculated for both the original and enhanced image and compared to assess how salience is affected by the distractor removal. A secondary objective is to evaluate the effect of distractor removal on object detection by feeding the enhanced images back as input to the CNN, creating a feedback loop. CNN's, salience models, and the human primary visual cortex consider basic features, such as edges, for their computations [7]. Since distractor removal reduces these features to decrease salience in the target region, it is interesting to investigate how it affects object detection. The inpainting methods are evaluated with the test set, ignoring images without benign objects, resulting in a set of 130 images in total.

Fig. 1. From left to right: original scan, detected benign and threat objects, and distractor removal applied with *Regionfill*.

IV. EXPERIMENTAL EVALUATION

All methods except *Uniform Inpainting* reduce the total visual clutter for at least 70% of the 130 images. *Regionfill* even achieves an efficiency of 95.4%.

Uniform Inpainting is the only method that fails to decrease distractors' salience and even increase it on average. The other methods successfully reduce the salience of distractors. *Regionfill* performs best by reducing salience in 96.9% of the 130 evaluated images by an average of 31.94%. These results are consistent with the clutter measurements. The salience of threat regions is maintained or even slightly increased by the distractor removal methods.

Since the primary goal of distractor removal methods is to mask distracting elements, inpainted benign items should no longer be detectable. Therefore, a crucial question is whether the detection model correctly rejects the removed benign items. *Regionfill* performs best with 86.51% correct rejections of the 583 inpainted benign items. This result, however, also means that at least 13.49% of the removed benign items are still detectable by the model. This can be partly explained by benign objects overlapping threats, as they cannot be fully inpainted. Otherwise, there is a risk that the threat will become unrecognizable. Moreover, semantic segmentation may fail when benign and threats are close together. *Uniform Inpainting* performs worst with a correct-rejection rate of 66.27%, presumably because the shapes of the inpainted items are still very prominent, which can be seen in the rightmost image in Fig. 3. All distractor removal methods lead to an increase in true-positive detections of benign items, indicating that the model can identify additional items after distractor removal is applied. On the original images, 583 of 1015 benign items could be detected successfully. After applying the methods to the images, the model detected further 1.4%-3.2% benign items, raising the true-positive rate for benign items from 58.4% to 59.8% - 61.6%. Furthermore, false-positive detections are decreased, at most by 28.87%.

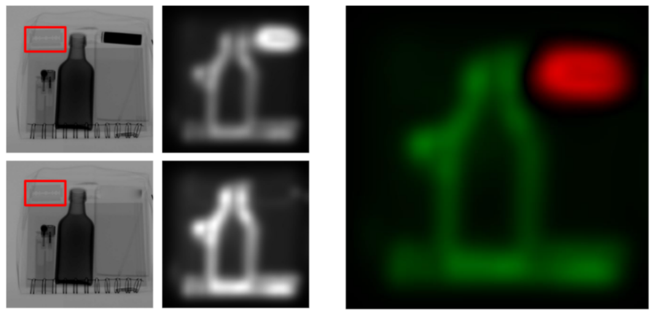

Fig. 2. From left to right: The original image with a marked threat and its enhanced version, their corresponding salience maps, and the comparison of the two maps. Red denotes a reduction in salience, and green an increase.

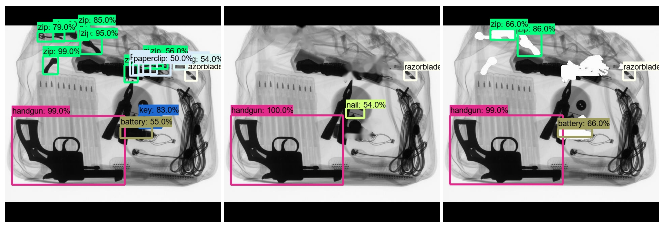

Fig. 3. Predictions on the original scan and after *Regionfill* and *Uniform Inpainting* are applied. *Regionfill* leads to a higher confidence and a new detection, while *Uniform Inpainting* fails to mask all distractors.

V. CONCLUSION

The experiments demonstrate that removing distracting items positively influences the scans' salience and successfully reduces visual clutter. The results suggest that detecting benign items in combination with distractor removal methods facilitates the visual search task, as clutter and salience are influential factors. This assumption is supported by the positive effect of the distractor removal on the object detection. Moreover, supplemental detection of benign objects provides additional information and is probably more practical than sole threat detection. While threats must not be overlooked, benign detection must only be accurate, without the need to detect all present benign objects. Furthermore, benign items are not usually deliberately concealed.

An essential disadvantage of distractor removal in 2D is that image information is artificially altered or removed without revealing new information to the viewer. Therefore, removing objects could be misleading. We hypothesize that this disadvantage is omitted as soon as more information about the bag is available, such as when working with computed tomography that provides volumetric data of the bag. Distractor removal techniques still have to be applied with caution. For example, removing objects that are part of more complex constructions can have an undesirable effect by making the entire construction unrecognizable.

References

[1] M. Baştan, "Multi-view object detection in dual-energy x-ray images," *Machine Vision and Applications*, vol. 26, no. 7, pp. 1045–1060, 2015.

[2] F. Bornemann and T. März, "Fast image inpainting based on coherence transport," *Journal of Mathematical Imaging and Vision*, vol. 28, no. 3, pp. 259–278, 2007.

[3] C. Chen, X. Du, L. Hou, J. Kim, P. Jin, J. Li, Y. Li, A. Rashwan, and H. Yu, "Tensorflow official model garden," 2020. [Online]. Available: https://github.com/tensorflow/models/tree/master/official

[4] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting," *IEEE Transactions on image processing*, vol. 13, no. 9, pp. 1200–1212, 2004.

[5] T. Franzel, U. Schmidt, and S. Roth, "Object detection in multi-view x-ray images," in *Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*. Springer, 2012, pp. 144–154.

[6] O. Fried, E. Shechtman, D. B. Goldman, and A. Finkelstein, "Finding distractors in images," in *Computer Vision and Pattern Recognition (CVPR)*, June 2015.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[8] D. K. Jain *et al.*, "An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery," *Pattern Recognition Letters*, vol. 120, pp. 112–119, 2019.

[9] R. T. Kneusel and M. C. Mozer, "Improving human-machine cooperative visual search with soft highlighting," *ACM Transactions on Applied Perception (TAP)*, vol. 15, no. 1, pp. 1–21, 2017.

[10] X. Liu, A. Gale, and T. Song, "Detection of terrorist threats in air passenger luggage: Expertise development," in *2007 41st Annual IEEE International Carnahan Conference on Security Technology*. IEEE, 2007, pp. 301–306.

[11] Q. Lu, *The utility of X-ray dual-energy transmission and scatter technologies for illicit material detection*. Virginia Polytechnic Institute and State University, 1999.

[12] V. A. Mateescu and I. V. Bajic, "Visual attention retargeting," *IEEE MultiMedia*, vol. 23, no. 1, pp. 82–91, 2015.

[13] MATLAB, *version 9.8.0.1380330 (R2020a)*. Natick, Massachusetts: The MathWorks Inc., 2020.

[14] J. S. McCarley, A. F. Kramer, C. D. Wickens, E. D. Vidoni, and W. R. Boot, "Visual skills in airport-security screening," *Psychological science*, vol. 15, no. 5, pp. 302–306, 2004.

[15] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar, H. Lobel, and M. Carrasco, "Gdxray: The database of x-ray images for nondestructive testing," *Journal of Nondestructive Evaluation*, vol. 34, no. 4, p. 42, 2015.

[16] S. Michel, S. Koller, M. Ruh, and A. Schwaninger, "The effect of image enhancement functions on x-ray detection performance," in *Proceedings of the 4th International Aviation Security Technology Symposium*, 11 2006.

[17] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[18] A. Schwaninger, S. Michel, and A. Bolfing, "Towards a model for estimating image difficulty in x-ray screening," in *Proceedings 39th Annual 2005 International Carnahan Conference on Security Technology*. IEEE, 2005, pp. 185–188.

[19] Y. Sterchi and A. Schwaninger, "A first simulation on optimizing eds for cabin baggage screening regarding throughput," in *2015 International Carnahan conference on security technology (ICCST)*. IEEE, 2015, pp. 55–60.

[20] G. Touya, B. Decherf, M. Lalanne, and M. Dumont, "Comparing image-based methods for assessing visual clutter in generalized maps," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 2, pp. 227–233, 2015.

[21] 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)*. IEEE, 2013, pp. 1140–1145.

[22] C. Wloka, T. Kunić, I. Kotseruba, R. Fahimi, N. Frosst, N. D. Bruce, and J. K. Tsotsos, "Smiler: Saliency model implementation library for experimental research," *arXiv preprint arXiv:1812.08848*, 2018.