

PRIP-TR-151

September 21, 2021

Benign Object Detection and Distractor Removal in 2D Baggage Scans

Anna Sebernegg

Abstract

X-ray screening significantly impacts security applications such as baggage handling to help detect objects, especially threats such as explosives or weapons, within closed luggage otherwise not visible to the naked eye. However, the generated X-ray images are challenging to interpret due to the targets' weak visual signals in high background noise levels and the compact assembly of rotated and superimposed objects in bags. The complexity of X-ray scans and the high intra-class variability of threats make appearance-based threat detection difficult for both human operators and automated systems. Consequently, generic appearance-based threat detection systems are hardly available in practice, and baggage screening still depends highly on human operators. Nevertheless, further developments of automatic baggage inspection are desirable to support the visual search task of screeners. This work proposes utilizing automatic benign object detection as a diagnostic aid, for instance, to remove distractors from the images through image inpainting. By reducing the number of distractive benign objects in the data, regions of interest could gain faster attention. The applied distractor removal methods successfully reduced visual saliency in regions of distractors and decreased the overall visual clutter of the X-ray scans.

1 Introduction

After 9/11 and various other incidents, security measures at airports and other public places have increased. Since then, more and more security research has been conducted to produce supporting technologies. X-ray screening and inspection systems were further developed to ensure that objects and defects can be detected non-destructively. Breakthroughs have been made, particularly in baggage screening, allowing to check every piece of luggage for prohibited items before a flight takes place. Nowadays, industrial X-ray screeners are widely used at security checkpoints [10]. Threat detection became one of their primary tasks alongside industrial quality controls, analysis of products, inspections of cargos, and archaeological discoveries [39].

The baggage screening process is increasingly automated, especially liquid and explosive detection systems for cabin baggage screening have evolved [59]. However, generic appearance-based threat detection systems are hardly available in airports [3, 41], assumingly due to the multiple challenges that come with the nature of baggage scans, the high intra-class variability and consequently the deficient detection accuracy. Therefore, current baggage scanners with automated subsystems still depend on human operators to perform various vigilance tasks that require sustained attention across extended periods [59, 67, 57]. These are visual search and decision-making tasks that demand human operators to remain attentive to prevent costly errors [21, 28]. Examples include inspecting X-ray images, alarm resolution, and monitoring the performance of automated systems. However, the desired ability of human operators to maintain the focus of cognitive activity on the given stimuli over prolonged periods is negatively affected by several factors. Human screeners often work under stressful environments due to factors like time pressure and high noise levels [37, 31]. At the same time, they have to detect weak and infrequent visual signals in a high background noise level. Visual search and object detection is additionally impaired by the unpredictable content and arrangement of objects in the bags and the nature of baggage scans, which can be difficult to interpret correctly [13]. These conditions can have a straining effect on the human operator's perceptual and cognitive abilities, leading to an increased error rate or a decrease in reaction rate over time [21, 37, 50].

To reduce the workload of the human operators, further developments of automatic baggage inspection, including appearance-based threat detection, are desirable as supportive systems for screeners [3]. However, as many different threat classes and schemes to conceal prohibited items exist, automatic and generic threat detection alone is not yet feasible. Consequently, designing more advanced algorithms and the automation of X-ray signal detection in baggage screening alone is insufficient to ensure performance at present [20, 31]. As the human element plays an important role, it is essential to improve human screener's capabilities [31].

One possible way to improve the accuracy and reliability of human performance in detecting threats could be to enhance X-ray scans by utilizing automatic object detection as a diagnostic aid. Like computer-aided detection systems (CAD) used in the medical field [27], detected regions of interest could be processed to focus the viewer's attention on critical content that requires further investigation. In medical radiography, CAD systems search for conspicuous areas (potential diseases) by applying pattern recognition and visually high-

lighting them to increase their saliency for human operators [27]. Similar techniques could be adapted for baggage screening to facilitate the visual search task of screeners. The visual salience of automatically detected threats, limited to easily identifiable shaped ones, could be enhanced by soft highlighting, described by Kneusel and Mozer [27]. Another potential application proposed in this work is to remove distractors from the X-ray images by inpainting detected benign objects that negatively contribute to visual clutter. By reducing the number of distractive objects (distractors) in the data, regions of interest could gain faster attention. Furthermore, the image augmentation could help draw the human operator’s attention to areas that include low-salient and potential unnoticed threat items. Distractive information could be filtered out by detecting a limited number of benign objects with distinctive features allowing a reliable detection. The proposed method aims to reduce the overall visual clutter of the X-ray scans and reduce the visual saliency in regions of detractors while ideally increasing it in the rest of the image.

1.1 Overview

The structure of the remainder of this bachelor thesis is as follows: Section 2 gives a short overview of the background of 2D baggage scans and the challenges they present for object detection due to their complex nature. Moreover, terms such as visual saliency are covered. Section 3 presents the aim of the work and explains the proposed methods in more detail. The design of carried out experiments and the structure of the database are discussed in Section 4. The quantitative evaluation can be found in Section 5, while the last section contains the conclusion and future research considerations.

2 Background and Related Work

2.1 Industrial 2D X-Ray Scans

In industrial radiography, 2D images are obtained by measuring the degree of absorption of the X-ray beam by the objects [26]. An X-ray source generates the beam, while an opposite receptor intercepts the rays that pass through the object, as illustrated in Figure 1 [68]. Since different materials absorb and scatter X-rays differently due to their different density and atomic properties, the internal structures of objects or luggage can be made visible [58]. Steel, for example, has a very high density. Therefore, only high-energy X-rays can pass through while lower-energy X-rays are absorbed. On the other hand, high- and low-energy X-rays can pass organic materials like bread because of their low density [58]. Unlike in medical radiography, the lower the density and thickness of a material, the brighter it is displayed on the industrial X-ray image [26, 63]. Higher density materials or deep objects are displayed dark, very high-density materials like lead crystal, cement, and different metals are visualized black. An example X-ray scan is given in Figure 2. The scanned case is of organic material and accordingly visualized very light. The alkaline batteries contained in the lower-right corner, on the other hand, have a high density and, therefore, colored black.

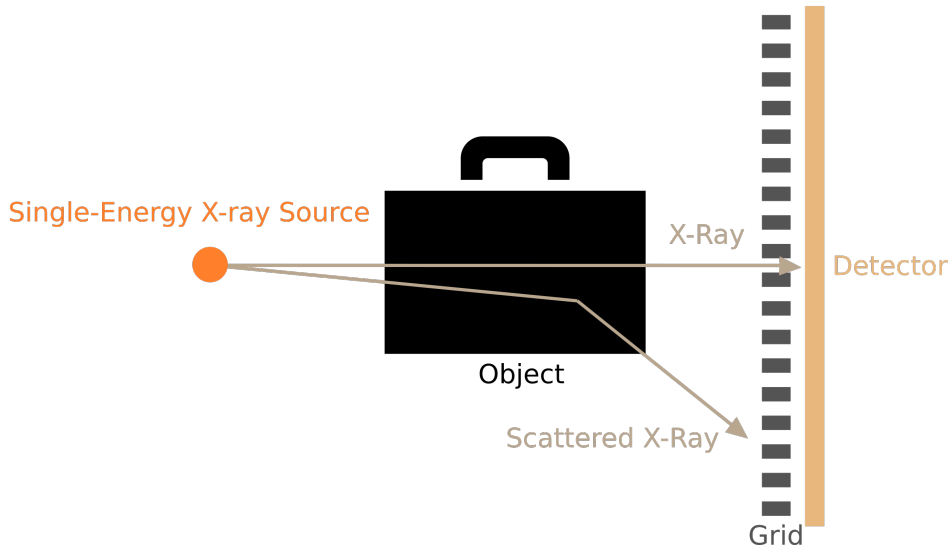


Figure 1: Simplified illustration of an Single-View baggage scanner.

2.2 Challenges for Object Detection in X-ray Scans

Due to the nature of 2D baggage scans, they cause several challenges for appearance-based object detection, both for human operators and for automatic threat detection [55, 65, 3, 33, 51].

X-ray images differ from conventional photographic images fundamentally in the information they provide [3]. Information in photographs – like **color**, **depth**, **texture**, and the **influence of light** on volumes such as reflections and refractions – is missing in X-ray

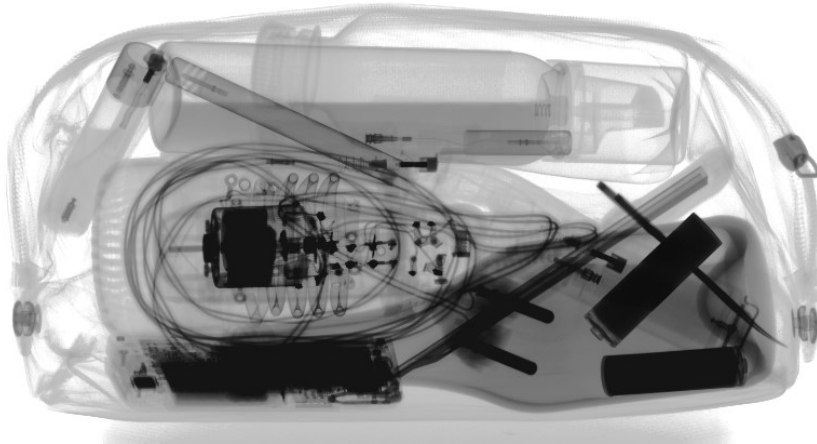


Figure 2: Example grayscale baggage scan.

images. This loss of familiar information leads to object representations that look very different from what people are used to. Consequently, human operators require extensive training and support to recognize various objects in baggage scans [45, 28]. On the other hand, X-ray images provide insights conventional photography is not able to produce. Since X-rays can pass through low densities, it is possible to see through otherwise opaque surfaces such as suitcases, enabling automatic threat detection at airports or medical examination of our skeleton [3]. The X-rays can pass through several objects before they are detected, introducing **transparency** to X-ray images where a single pixel can contain information about several overlapping objects [13]. A challenge introduced by the transparent overlaps is the segmentation of different object-information from one another [13]. The transparency property can be seen in Figure 2, where both the folds of the scanned case and the objects behind it are visible. High-density objects absorb the vast majority of X-rays. Therefore, they are opaque on X-ray images and **obscure overlapping or enclosed** objects, while low-density objects are almost invisible [3, 42].

Another challenging factor is that objects within luggage occur at any **orientation** in- and out-of-plane rotations [14, 3]. Especially out-of-plane rotations as visualized in Figure 3 lead to difficulties in recognizing objects correctly.

Tightly packed luggage can lead to a high degree of **clutter** and overlaps, severely increasing the **complexity** of the baggage scans [65]. Furthermore, in the appearance of objects, both in shape and density, a broad **intra-class variability** exists [51]. A single category, such as knives, is made up of various items manufactured by multiple brands with different characteristics, such as the materials used to make them. In addition, the different categories that are considered threats or prohibited at airports are vast.

Finally, a challenge that primarily affects categories of prohibited items are **schemes to conceal** objects [33].

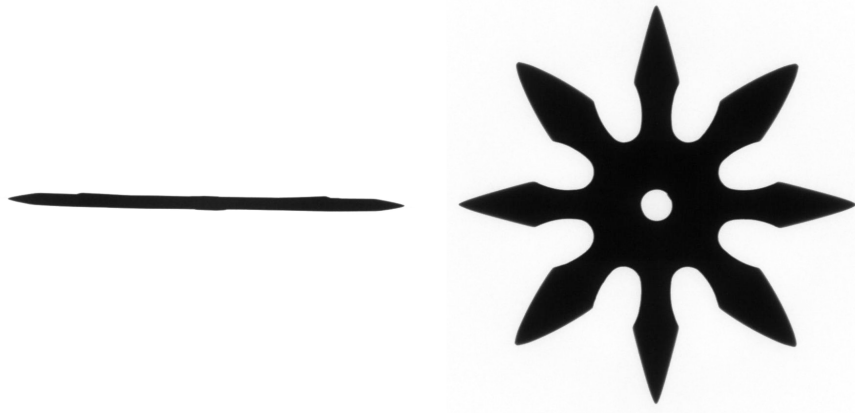


Figure 3: Shuriken scanned from two different view points with out-of-plane rotation.

2.3 Visual Search and Saliency

Visual search is a perceptual task where a target object has to be found amid other objects, usually referred to as distractors [6]. In baggage screening, human operators have to perform visual search tasks excessively.

Target saliency is a factor that is known to influence the accuracy and speed of the conducted visual search [6]. Saliency describes how prominent an item or image region is when surrounded by neighboring objects [6]. Items referred to as high-salient immediately catch the eye, while those that do not stand out are low-salient. An object’s saliency depends on visual features such as color, size, orientation, luminance, motion, and its neighborhood.

2.4 Related Work

Extensive research is being conducted in appearance-based threat detection, using both machine learning and deep learning methods [1, 24]. Some work also addresses the problem of material discrimination in X-ray images using material classification [4]. However, at the time of writing, no relevant work could be found where benign object detection in baggage scans is the main objective.

Image processing is a broad field utilized in baggage screening to improve readability for human operators and automatic detection systems [1, 44]. Edge enhancement, pseudo coloring, or noise reduction are some examples that are investigated for baggage scanning. The image processing techniques that are probably most similar to the distractor removal method used in this work are material filters, where parts consisting of certain materials or densities are filtered. One example is the organic-only filter mentioned by Michel et al. [44] and Schwaninger et al. [53], in which only organic materials are displayed. Saliency-driven image manipulation techniques such as distractor removal or attention retargeting are especially explored for photography and are less common in baggage screening [34, 38].

3 Methodology

Current systems for scanning carry-on luggage require human operators to visually search for threat items in X-ray scans and perform alarm resolution [31, 21]. However, baggage scans are very complex by nature and contain high background noise levels [65]. One possible way to improve the accuracy and reliability of threat detection in such scans could be to utilize automatic object detection as a diagnostic aid. Thereby, the scans could be enhanced to focus the viewer’s attention on critical content that requires further investigation.

This thesis’s main objective is to use the results of automatic object detection to enhance regions of 2D baggage scans in a way that reduces visual clutter and focuses attention on regions that potentially contain threats. To achieve this goal, two different image enhancements are performed based on object detection, where each of the methods modifies the visual salience in the corresponding region. The first method highlights detected threats by a unique visual feature to allow faster alarm resolution. The second method applies distractor removal to the 2D baggage scans to mask out detected benign objects that negatively contribute to visual clutter. Therefore, this thesis’s main aspects are selecting and training an appropriate object detection model to detect both threat and benign objects in 2D baggage scans and to study the effects of the applied image enhancement methods on human visual attention. The image enhancement methods will be evaluated by utilizing the Itti-Koch-Niebur Saliency Model (IKN) [23] and Quad-tree-clutter [25].

A secondary objective is to evaluate the effect of distractor removal on the object detection model by feeding the enhanced images back as input to the Convolutional Neural Network (CNN), creating a feedback loop. CNNs, saliency models, and the human primary visual cortex consider basic features, such as edges, for their computations [17, 29, 23]. Since distractor removal reduces such basic features to decrease salience in the target region, it is interesting to investigate how this processing affects the overall object classification through the trained CNN. At least, it is expected that the removed object will no longer be detected. If the image enhancement also positively affects the overall salience of the image, it could positively impact object classification.

The content of this thesis is summarized as follows:

1. Discussing the composition of the database used in this work
2. Applying transfer learning on a pre-trained model to obtain the object detection model
3. Approximating the foreground regions of the detected objects
4. Performing two different image enhancements on the calculated foreground regions:
 - (a) highlighting detected threats by a color overlay
 - (b) removing detected benign objects from the image by inpainting
5. Evaluating the image enhancements
 - (a) effects on human visual attention

(b) effects on the object detection model

3.1 Dataset

The primary dataset used in this thesis is obtained from the public mono-energy X-ray database *GDXray* created by Mery et al. [40]. This database can be used free of charge, but for research and educational purposes only. It contains scans of various hand luggage such as backpacks, wallets, pencil cases, and single images of threat items such as knives, handguns, razor blades, and shuriken [20]. Example baggage scans for this database are given in Figure 4.

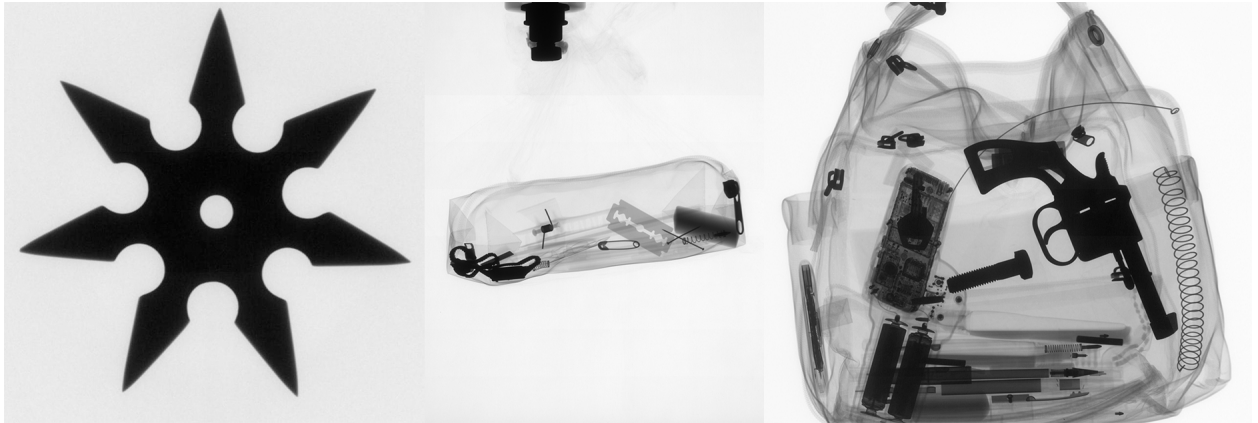


Figure 4: Example images obtained from the GDXray database (Baggage)

An additional, smaller dataset was generated to extend the database with baggage scans that contain a high degree of visual clutter and additional object categories. This dataset was created through cooperation with the CT Research Group at Campus Wels in upper Austria using the dual-source computed tomography scanner RayScan 250E¹. Example scans are given in Figure 5.

All twelve objects of interest (shuriken, handgun, knife, razor blade, zip, nail, staple, key, battery, paperclip, spring, and screw) are labeled with oriented rectangles. The exact structure of the dataset and the annotation process is described in Section 4.1.

Without cooperating institutions, it is difficult to access data for a security domain like baggage screening. Most of the inspection systems and training software are proprietary. Moreover, public databases are very limited [41]. From the literature used in this thesis, only three image databases are publicly available without additional restrictions such as a Non-Disclosure Agreement (NDA) or payment: GDXray, SIXray, and COMPASS-XP Dataset [40, 43, 18]. The only reasonably large volumetric baggage dataset found at the time of writing is the ALERT CT dataset from Northeastern University². However, this dataset

¹Research Group Computed Tomography: <http://www.3dct.at/cms2/index.php/en/> I am very grateful to Dr. Christoph Heinzl and Johanna Herr who made the cooperation possible.

²<http://www.northeastern.edu/alert/transitioning-technology/alert-datasets/>



Figure 5: Example images from the dataset generated in cooperation with the CT Research Group at Wels Campus in upper Austria.

could not be used in this thesis because it is only available under an NDA where all generated publications must be reviewed and approved by Northeastern University/ALERT Research Evaluation Advisory Panel (REAP). The absent variety of public databases is a limiting factor for research in security domains and has also influenced this work. Due to the missing public volumetric datasets at the time of writing, the primary approach to remove volumes of benign objects from baggage scans by applying an opacity transfer function to reduce the number of distracting items in 3D baggage scans was approximated and investigated solely with 2D data. Furthermore, the focus on the existing 2D databases lies on prohibited items and not on everyday objects. Therefore, only a handful of benign object categories are selected as objects of interest in this work.

3.2 Object Detection Model

For applying the two image enhancement methods, in which particular objects are either removed or highlighted, the objects of interest must be located and classified in advance by a suitable object detection model. The final object detector should detect twelve different objects, of which four can be classified as threat items while the remaining eight objects are benign. The results are then given as input to the image enhancement methods. It is important to note that regions of interest are required for the image enhancement methods. Therefore, once the objects are detected, semantic segmentation is performed in MATLAB to identify the masks of the classified objects. The given bounding boxes from the object detector simplify this segmentation. Instead of doing the object detection and segmentation separately, another option would be to use an instance segmentation model such as Mask R-CNN.

The object detection model used in this thesis is received by applying transfer learning to a pre-trained EfficientDet model (D1) [8]. The model is pre-trained on the COCO 2017 dataset [30] and is provided by the TensorFlow Object Detection API, which is "an open-

source framework built on top of TensorFlow that makes it easy to construct, train, and deploy object detection models" [8]. The training process itself is done with TensorFlow 2 on a local GPU.

3.2.1 Model Selection

Traditional machine learning approaches for object detection are built on handcrafted features, which are manually designed by experts instead of learned from data [70]. Therefore, it is necessary to transform the raw data (such as pixel values) into a suitable representation or feature vector before using it as input to a classifier such as a Support Vector Machine (SVM) [70]. Well-known feature detectors are, for example, Scale-Invariant Feature Transform (SIFT) [32] and Histogram of Oriented Gradients (HOG) [12]. Due to the various properties and appearance of different objects, backgrounds, and illumination conditions in scenes, it is not easy to design a well-performing feature descriptor that can be applied to multiclass object detection [70, 62].

Deep learning approaches such as Convolutional Neural Network (CNN), on the other hand, utilize representation learning, which means that they can work with the raw data and automatically learn the features needed for the following prediction task [19, 29]. Convolutional Neural Networks are state-of-the-art approaches for object classification tasks on images [46]. However, CNNs alone do not perform the additional localization task needed to retrieve the object position [62], which is necessary to identify and enhance the local region of the detected objects. Widely applied architectures that utilize both classification and regression for object detection are two-stage detectors such as Region-based CNN (R-CNN) and its variants [70], which combine region proposals with a CNN as a feature detector [16]. Since the proposal of R-CNNs, many other object detection models based on CNNs have been suggested [70]. A recently introduced object detector, which mostly follows the one-stage detector paradigm, is EfficientDet [61].

For this work, an object detector based on a CNN is chosen because they can achieve remarkable accuracy on object detection while often requiring less expert analysis and fine-tuning for the feature detection than traditional approaches [48]. Moreover, CNN-based object detection models can be re-trained using a custom dataset [48]. Therefore, a single framework can be applied to multiple domains.

A common reason for choosing traditional methods is, for example, when the required quantity of data for training is unavailable [48]. Even though the GDXray database contains only 8150 baggage scans, deep learning is already successfully applied to it multiple times [41, 24, 2]. Therefore, the size of the used dataset should not be an issue.

In this work, EfficientDet (D1) is chosen as a model. Compared to other models such as RetinaNet, YOLOv3 or Faster R-CNN [61], it achieves similar or even higher accuracy on Microsofts Common Objects in Context (COCO) dataset [61]. At the same time, it is much smaller and more efficient compared to other state-of-the-art detection models [61]. Furthermore, it is available as pre-trained model in the "TensorFlow 2 Detection Model Zoo" [8].

3.2.2 Training of the Model

As training a deep neural network requires a large dataset, the training process is often costly in time and other resources [48]. A common technique to reduce training time is transfer learning, where information from previously learned tasks is reused for the learning of new related tasks [56].

In this work, transfer learning is applied to the EfficientDet model (D1) provided by the TensorFlow Object Detection API [8]. The provided model is pre-trained on the COCO 2017 dataset [30]. The COCO dataset provides images of 1.5 million common object instances; all non-Xray images [30]. Therefore, transfer learning from non-X-ray to X-ray images is applied. A similar application for transfer learning is already tested, for example by Hoo-Chang Shin et al. [56], where they applied transfer learning on ImageNet [52] for medical image recognition tasks.

3.2.3 Evaluation of the Model

The final object detector should be able to detect four threat- and eight benign-items. The performance for each object is evaluated by comparing the False Negative Rate or Miss Rate (FNR), False Positive Rate or Fall Out (FPR), True Positive Rate or Recall (TPR), Precision (P), and Accuracy (A). The performance of the object detection model is visualized in the form of a confusion matrix. Additionally, the performance of the two main categories (threat and benign items) is discussed. This work’s main focus lies in detecting benign objects, as they are later removed in the image enhancement step to reduce the background noise level. Therefore, it is desirable to correctly detect as many benign objects as possible. Furthermore, it is desired that no threat item is misclassified as benign by the neural network, as it would then be removed by the image enhancement. Falsely removing a threat item would fail to focus the viewer’s attention on critical content and hide essential information.

Unfortunately, a fair comparison to other works using the GDXray is hardly possible: Training and testing sets are different between publications, and the results are presented using several metrics [41]. Furthermore, a custom dataset is used in combination with the GDXray database making comparisons even harder.

3.3 Image Enhancement

Image enhancement techniques improve the information content and quality of an image [49]. One of their desired results is to magnify important details in images that might otherwise not be immediately visible in order to create image features that are more apparent to detect for both human and automatic image analysis [60, 49].

As described in Section 2, human operators and automatic object detection algorithms must locate and classify rare and weak visual signals in baggage scans that contain high background noise levels [37, 31]. Image enhancement methods could be one way to improve the accuracy and reliability of both human and algorithmic threat detection performance on this challenging data. By amplifying specific visual signals, the viewer’s attention could be directed to critical content that requires further investigation.

To draw attention to already detected objects, highlighting techniques can be used. This approach is, for example, utilized in Computer-Aided Detection (CAD) where potential diseases in medical images are marked [27]. CAD apply automatic pattern recognition and classifiers to find the regions of interest (ROI) and highlight them visually to increase their saliency for human operators [27]. Similar techniques could be adapted for baggage screening to facilitate screeners’ visual search tasks and automatic detection systems. However, critical regions in baggage scans are not only areas where threat items are easily recognizable, but those where many objects are clustered together or where objects are not easily recognizable. Inconspicuous regions may as well contain threat items. A possible enhancement of such regions is to reduce the overall visual background noise, for example, by removing distractors from the image. Distractors are visually distracting items, i.e., regions that divert attention from the main subjects and thus diminishing the overall image quality [15]. Distractor removal is, for example, applicable in photography by using inpainting methods such as Photoshop’s Content-Aware Fill, which tries to reconstruct and fill in the background of a removed object, to improve the composition of an image [15]. A similar automated approach can remove distractors, such as everyday objects, from baggage scans. By removing the distractors, the overall visual clutter can be reduced and may focus the viewer’s attention on critical objects.

In this paper, two different types of image enhancement methods are implemented and evaluated on 2D baggage scans: The first method highlights potential threat items detected by the object detection model with a particular confidence score. The second method inpainted regions of everyday objects to remove distractors from the baggage scan and to further reduce the visual clutter. For both enhancements, the detection model results are used to calculate a more specific region of interest. The implementation of the image enhancement methods is done in Matlab R2020a combined with the Image Processing Toolkit.

3.3.1 Automatic Distractor Removal

Distractor removal removes distracting or irrelevant information from the data to improve the signal to noise ratio [15]. What information is considered irrelevant depends on the context and often is selected by hand. In the context of this work, automatically detected benign objects are treated as distractors and therefore accounted as irrelevant. This type of information filtering aims to reduce the image’s visual clutter and draws attention to the main subjects [15].

In this thesis, automatic distractor removal is evaluated on 2D baggage scans. The goal is to reduce the amount of distracting benign objects in the baggage scans to diminish clutter and shift the saliency to other image regions to gain potentially faster visual attention on regions of interest. Thereby, the method aims to decrease salience in the filtered regions while maintaining or even increasing salience in the rest of the image, especially in threat item regions. When applying the method, the salience of unprocessed regions should not be influenced negatively, i.e., the salience of unfiltered image regions should not decrease in relation to the overall image salience. Whether an image region stands out or not usually depends on local properties such as contrast, brightness, color, size, and other image features such as if edges are present or not [31]. Filtered regions should therefore reduce such salient

features and be as uniformly looking as possible.

The best result for distractor removal in baggage scans is likely to be achieved with 3D data, as they contain more information about the baggage contents than projected 2D data. In 3D scans, distractors could be removed from the volume without affecting neighboring objects in front, behind, or inside the target, for example, by applying customized opacity transfer functions. Transfer functions used in direct volume rendering assign optical properties such as color and opacity to data values [36]. Voxels of segmented distractors could be set to an opacity of 100%, making them invisible in the volume rendering. On the other hand, distractor removal in 2D baggage scans is very limited. As luggage itself is three-dimensional, the projected 2D scans contain less information than the original baggage content, especially concerning objects that lie behind denser items. However, since hardly any suitable databases of 3D baggage scans are publicly available, only 2D data is used in this work [41]. One of the main drawbacks of applying distractor removal to 2D data is that overlapping objects are not preserved in projections. Therefore, the distractor removal techniques cannot recover or use the information of overlapping objects, and the implementation must take into account regions where threat and benign items overlap. In addition, the information used to replace the filtered regions must be carefully selected to avoid adding further distractions or noise to the image, for example, by selecting only background regions.

A potential solution for distractor removal on 2D baggage scans is to inpaint the region of interest with low-salient background regions, e.g., areas containing no object except the bag itself. A possible result of a distractor removal method using inpainting is given in Figure 6. Inpainting methods modify local areas of an image by filling them with image-specific information such as neighboring pixel values and therefore replacing the regions information [5, 15]. They are ordinarily used to restore paintings or remove objects from photographs [5]. It is important to note that inpainting techniques cannot fully restore the processed region’s information. Instead, they try to approximate it by using the limited information of the remaining image.



Figure 6: Image enhancement by inpainting. Left: original image; Middle: bounding boxes detected by trained object detector; Right: detected benign items (e.g. springs) are removed with inpainting based on MATLAB’s regionfill method.

Inpainting methods for distractor removal are evaluated in two ways: First, the enhanced

baggage scan’s overall visual clutter is measured and compared to the original image’s visual clutter to determine if clutter is successfully reduced. For this global evaluation, the quad-tree-clutter measure for visual clutter proposed by Jégou and Deblonde [25, 64] is applied. The required Quadtree Decomposition is done through MATLAB’s function `qtdecomp(I,threshold)` [35]. Secondly, local saliency changes are measured using the Itti-Koch-Niebur Saliency Model (IKN) [23] provided by the *Saliency Model Implementation Library for Experimental Research* (SMILER) [69]. A saliency map is calculated for both the original and enhanced image to assess whether salience is reduced in filtered regions and increased in others. The salience within the local regions of interest is then compared.

3.3.2 Highlighting

Highlighting, in the context of images, means drawing attention to an image region by making it visually prominent, similar to marking a text section with a highlighter. A target can be made visually prominent by artificially emphasizing it with a unique, salient visual cue such as color, orientation, or size [9]. Highlighting can effectively draw attention to the target and produce efficient search performance regardless of the number of objects displayed [9]. Current explosive detection systems for carry-on baggage screening (EDSCB) can already provide diagnostic aid by marking areas that may contain explosive material [54, 22]. A similar image enhancement technique could be applied to systems that perform threat detection to provide a faster focus on regions that might contain threat items so that alarm resolution can be performed faster.

In this work, a simple highlighting technique for threat detection is implemented. As the dataset used in this work contains only grayscale images, highly saturated colors can be used as a unique visual feature. An example of the highlighting is given in Figure 7.

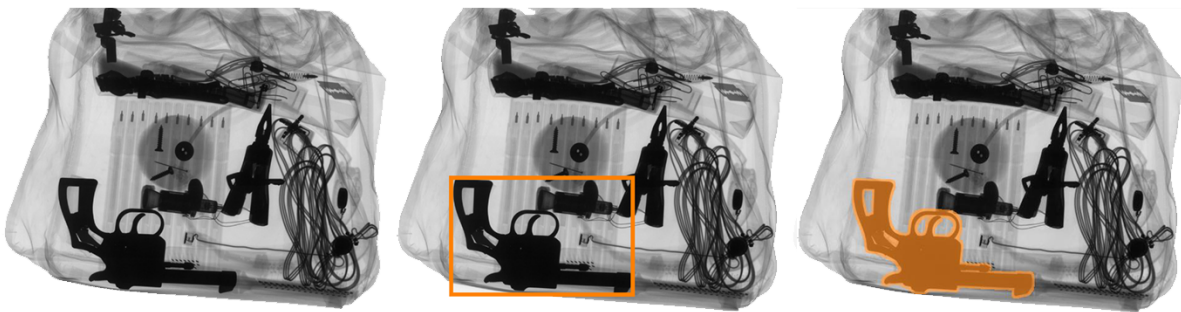


Figure 7: Image enhancement by highlighting. Left: original image; Middle: bounding boxes detected by trained object detector; Right: threat item is highlighted by color overlay.

3.3.3 Image Enhancement in a Feedback Loop

CNNs learn, among others, basic features such as edges, which are also used by the human primary visual cortex and saliency maps like IKN [17, 29, 23]. Since distractor removal discards basic features from the processed image region, this image enhancement method

could, to some extent, affect both human salience and CNNs object classification performance. Whether and how distractor removal influences object classification performance is evaluated by feeding the enhanced images back into the neural network as input, creating a feedback loop visualized in Figure 8. The results are then compared to the performance on the original test set. The main metrics used for the comparison are the confidence score and the number of correct and incorrect detected objects.

The evaluation process by creating a feedback loop consists of the following steps:

1. The unprocessed test set is used to evaluate the performance of the trained EfficientDet and forms the basis against which other results can be compared.
2. The resulting bounding boxes and object detection confidence scores are then used for distractor removal.
3. The enhanced images are routed back as inputs to the EfficientDet, and its results are compared to the performance on the original test set.

The last two steps could be repeated several times, but are only performed once in this work.



Figure 8: After enhancing the images based on the detected bounding boxes and confidence scores the processed images are fed back to the EfficientDet.

4 Experiments

4.1 Database

The database used in this thesis consists of 640 x 640 grayscale images obtained from the public mono-energy X-ray database *GDXray* and of baggage scans created in cooperation with the CT Research Group at Wels Campus in upper Austria.

This work focuses on detecting twelve different categories of objects of which silhouettes are given in Figure 9 and 10. The selected twelve objects can be further grouped into *threat items* and *benign items*. In this work, an object is categorized as threat if it is typically prohibited in hand luggage at airports, otherwise it is categorized as benign. The threat items of interest are *shuriken*, *handguns*, *knives*, and *razor blades*, while the benign items of interest are *keys*, *nails*, *zips*, *staples*, *batteries*, *paperclips*, *springs*, and *screws*. All twelve objects named above have in common that they mainly consist of inorganic materials, particularly metals. Therefore they are displayed very dark on X-ray images and have very little ‘transparency’. As double edge razor blades are very thin, usually with a thickness of under 0.25 mm, they are displayed way brighter than the other, thicker objects of interest. This property had to be considered, especially in calculating the regions of interest in both the labeling and image enhancement process.

The database is parted into three primary datasets (*Train*, *Validation* and *Test*), which are used for the training and evaluation of the object detection. A more precise breakdown of the datasets can be viewed in Table 1.

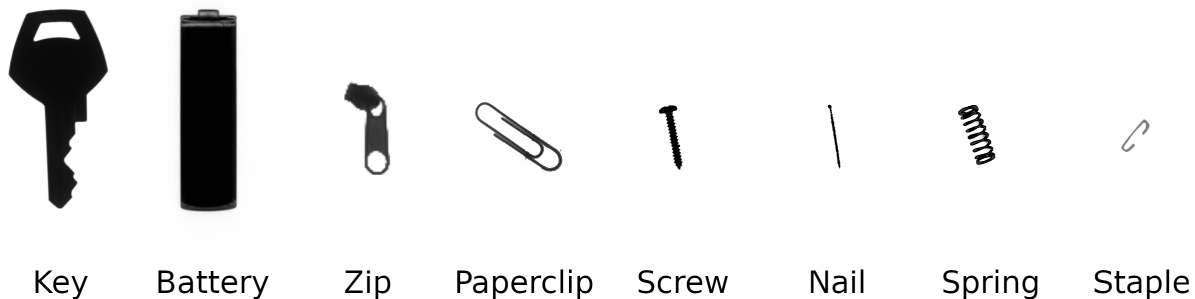


Figure 9: Silhouettes from the eight different types of everyday objects.

4.1.1 Data Acquisition

In cooperation with the CT Research Group at Wels Campus in upper Austria, 56 X-ray images of objects and cluttered bags were generated. The images were scanned with the dual-source computed tomography system *RayScan 250E*, which produces 2048 x 2048 resolution scans³. At least two views of each object/bag were recorded, and the resulting 2D

³<https://3dct.at/cms2/index.php/de/ausstattung/23-rayscan>

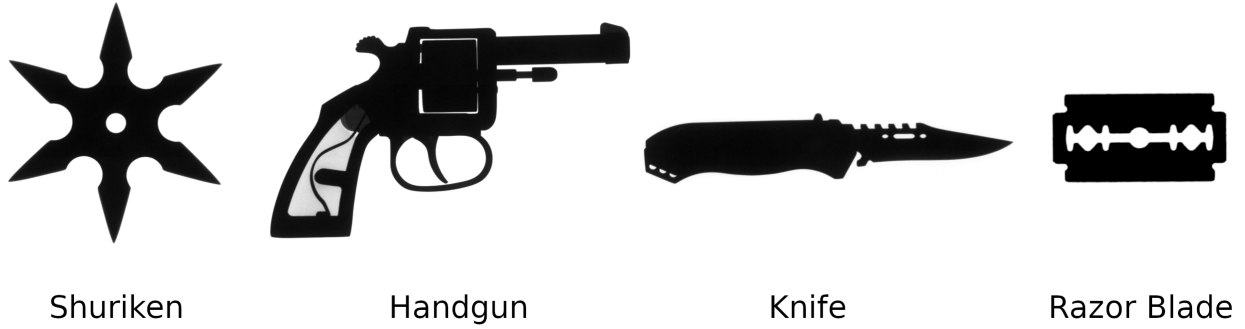


Figure 10: Silhouettes from the four different types of threat items.

scans contain both benign- and threat- objects such as lighters, spray cans, pepper spray, razor blades, batteries, zips, knives, and bottles. The images were edited by adjusting the brightness and contrast to enlarge the dataset by additional 83 images. Additional images were created by extracting objects of interest, such as batteries and screws, from the X-ray scans, creating single object images and combined images.

The remaining part of the dataset is obtained from the public GDXray database for X-ray testing and Computer Vision created by Mery et al. [40]. The subgroup "Baggage" used in this work contains a total of 8150 X-ray scans, of which only a subset of images is used. The scans are of various hand luggage such as backpacks, wallets, pencil cases, and single images of threat objects such as knives, handguns, razor blades, and shuriken.

Most of the images were already annotated with oriented rectangles [40]. However, some of the twelve objects of interest for this work were not. Therefore, all X-ray images used in this work were labeled from scratch with horizontal rectangles for consistency.

Overall, 3721 X-ray images were edited, labeled, and used in this work. A more precise breakdown of the total dataset can be viewed in Table 1.

4.1.2 Labeling

The labeling of the twelve objects of interest mainly was done by hand with the open-source, graphical image annotation tool `LabelImg` [66].

However, since a large portion of the dataset consists of single object recordings where the single object of interest is located in the center of the image, most of these images were labeled using a short MATLAB script that identifies the ROI as the area closest to the center of the image. The part of the script that segments the ROI is given in Listing 1.

In Figure 11, the script's steps are visualized: First, by calling the function `multithresh`, five threshold levels of the input image are calculated by utilizing MATLAB's implementation of Otsu's method [47, 35]. The thresholds are illustrated in the first column of Figure 11 by using the functions `imquantize` and `label2rgb`. The calculated thresholds are then used to

Table 1: Occurrence of objects of interest in the datasets

Object	Train		Validate		Test	
	Occurrence	Images	Occurrence	Images	Occurrence	Images
Knife	1599	1561	352	344	37	24
Handgun	433	416	84	82	66	65
Shuriken	528	504	116	110	38	31
Razor Blade	463	434	99	91	86	72
Zip	1285	332	254	73	310	78
Key	195	141	57	36	49	32
Nail	258	158	73	44	53	33
Screw	651	325	129	63	126	63
Spring	842	481	215	117	167	97
Staple	509	113	98	24	64	13
Battery	624	231	140	54	183	55
Paperclip	317	136	68	29	69	28
Total Images		2918		632		151

binarize the input image, as given in line 12 of Listing 1. For most of the input images, the fourth threshold value is selected. This selection is based on the assumption that the objects of interest are darker than the background since they consist of metal. Exceptions to this assumption are some images of razor blades. Although razor blades are also made of metal, their thinness makes them display brighter on X-ray images than the other objects of interest. The resulting binary image is further edited by applying morphological opening (erosion followed by dilation) and area opening to remove small connected components. The resulting binary image and the centroids of all remaining connected components are visualized in the second column of Figure 11. The image center is plotted as well. Next, the ROI is selected by the shortest distance between the image center and the region centroids. The final ROI is shown in the third column of Figure 11. Finally, the bounding box is calculated with the help of the function `regionprops`, which is from the Image Processing Toolbox and measures properties of image regions [35].

Listing 1: Matlab code sample: find ROI nearest image center

```

1 I_original = imread(path);
2 I = im2double(I_original);
3 [rows, columns, channels] = size(I);
4 center = [columns, rows] * 0.5; % image center
5
6 if channels > 1
7     I = rgb2gray(I);
8 end
9 I = imadjust(I);

```

```

10
11 threshold = multithresh(I, 5); % uses Otsu's method
12 BW = imcomplement(imbinarize(I, threshold(4)));
13 BW = bwmorph(BW, 'open');
14 BW = bwareaopen(BW, 20, 4);
15
16 stats = regionprops(BW, I, {'Centroid', 'PixelIdxList'});
17 centroids = cat(1, stats.Centroid);
18 distances = vecnorm((center - centroids)');
19 index = find(distances ~= min(distances)); % centered object
20 if ~isempty(index)
21     BW(extractfield(stats(index), 'PixelIdxList')) = 0;
22 end
23
24 BW = imdilate(BW, strel('disk', 4));
25 stats = regionprops(BW, I, {'BoundingBox'});

```

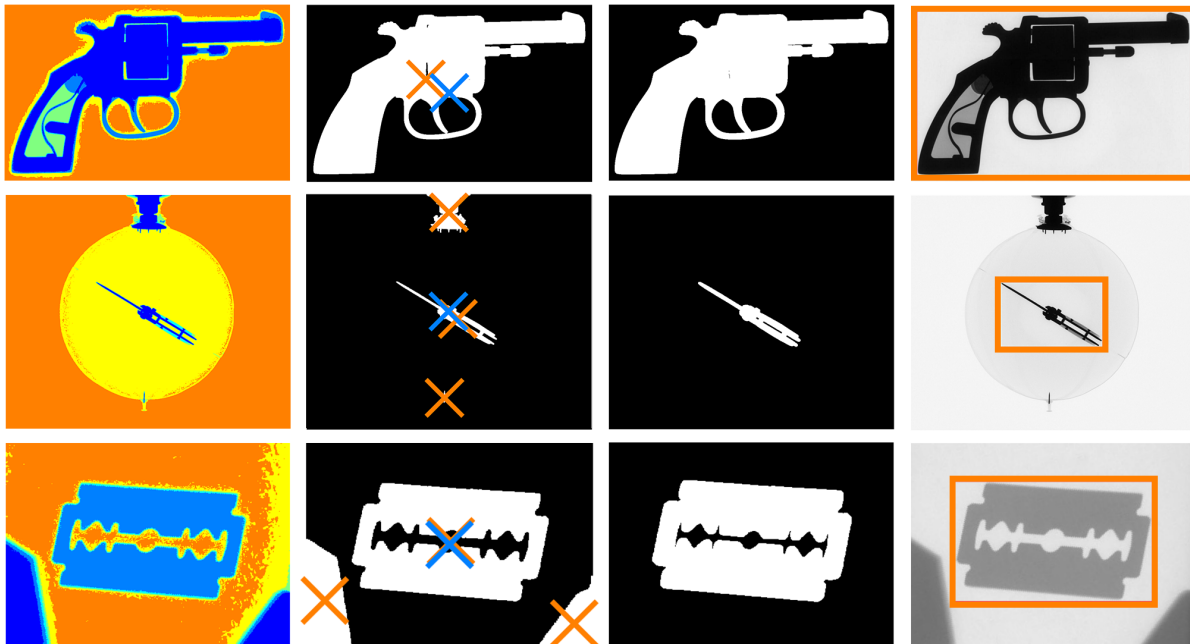


Figure 11: Visualizations for Listening 1. First column: quantized image using multithresh; Second column: binarized image, image center and centroids of regions; Third column: calculated ROI; Fourth column: bounding box.

4.2 Overview of Distractor Removal Methods

The chosen distractor removal 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.

To find a suitable inpainting method, several approaches are experimented with. This Section gives a short overview of the different approaches. All methods are implemented in MATLAB and are mostly based on provided algorithms from version R2019b [35]. The resulting distractor removal methods are evaluated and compared in Section 5.2.

4.2.1 Inpaint with Uniform Color

A potential simple approach to remove distractors from 2D baggage scans is to colorize the region of interest with a uniform color that blends in with the rest of the image. We experimented with the following three ways to extract the color from the given input image:

Inpaint with Max-Background-Color: The brightest grey value from the background of the image is selected and used to overwrite the pixel values of the ROI. For this purpose, the background region of the image must be identified. An image and its extracted background region displayed as binary mask are given in Figure 12.

Inpaint with Mean-Image-Color: The average grey value of the image is selected and used to overwrite the pixel values of the ROI.

Inpaint with ROI-Background-Color: A bright grey value within the region of interest is selected and used to overwrite the pixel values of the ROI. The following MATLAB code is used to select the color value:

```
thresh = multithresh(I(ROI_dilated), 5);
background_color = (thresh(5) + thresh(4)) * 0.5;
```

Example inpainted images with the extracted colors are given in Figure 13. Additional filter methods can be applied after inpainting to reduce edges or artifacts on the border of the colored area (e.g., Gaussian Filter or circular averaging filter), for an example see Figure 14.

4.2.2 Inpaint Coherent

`Inpaint Coherent` is a function provided by MATLAB that was introduced in R2019b [35]. It provides coherence transport based image inpainting as described by Bornemann and März [7].

`Inpaint Coherent` is called in the following way:

```
I_enhanced = inpaintCoherent(I_grayscale, ROI);
```

Example images inpainted with this method are given in Figure 15.



Figure 12: Image and its Binary mask for background region.

4.2.3 Inpaint Exemplar

`Inpaint Exemplar` is a function provided by MATLAB that was introduced in R2019b [35]. It provides exemplar-based image inpainting as described by Criminisi et. al [11].

`Inpaint Exemplar` is called in the following way:

```
I_enhanced = inpaintExemplar(I_grayscale, ROI, 'FillOrder', ...
                             'tensor', 'PatchSize', 14);
```

Example images inpainted with this method are given in Figure 16.

4.2.4 Regionfill

`Regionfill` is a function provided by MATLAB that was introduced in R2015a [35]. It inpaints a given image region using inward interpolation from the pixel values at the outer boundary of the area [35].

`Regionfill` is called in the following way:

```
I_enhanced = regionfill(I_grayscale, ROI);
```

Example images inpainted with this method are given in Figure 17.

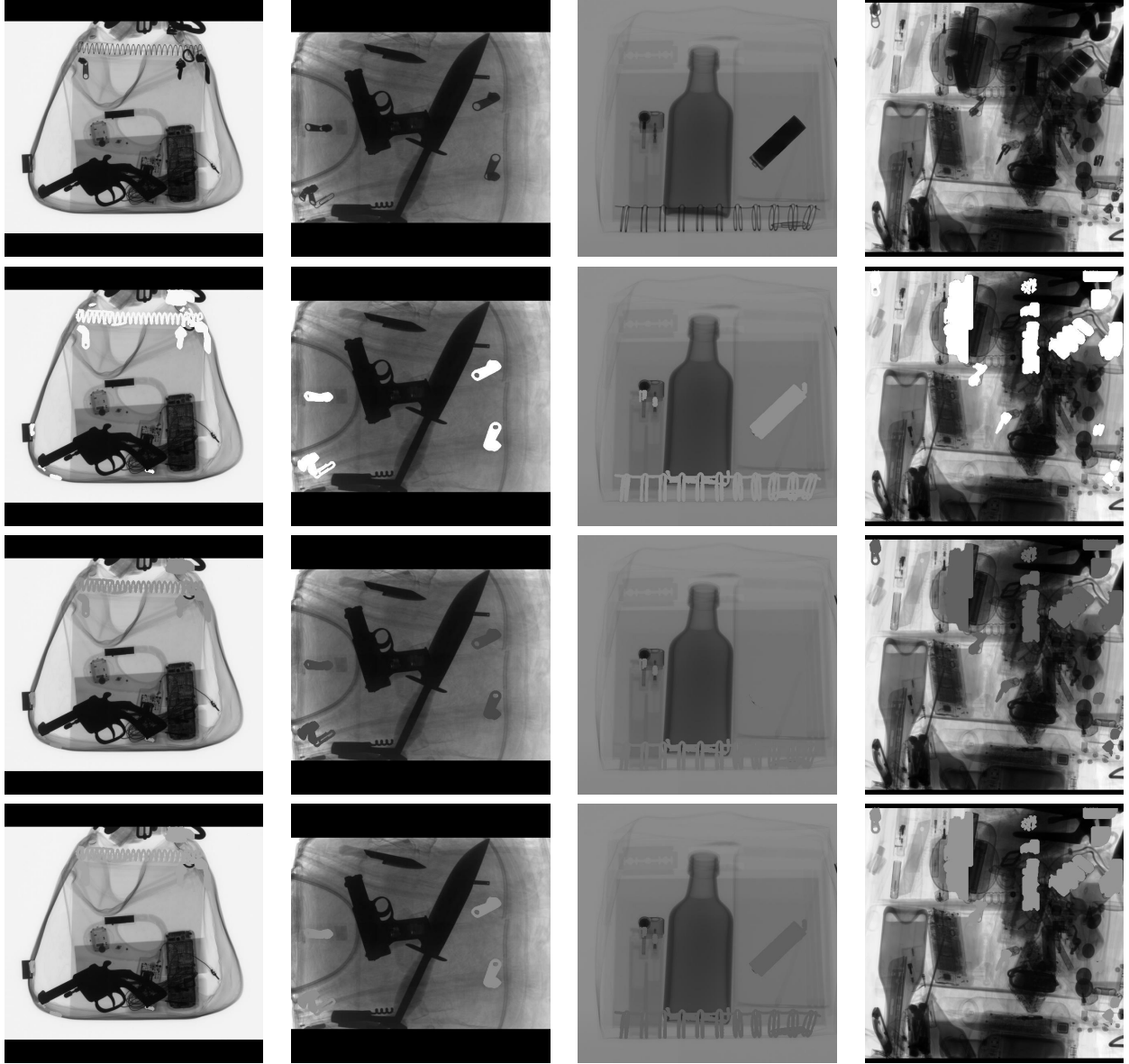


Figure 13: Inpaint with uniform color. First row: original images; Second row: images *Inpaint with Max-Background-Color*; Third row: images *Inpaint with Mean-Image-Color*; Fourth row: images *Inpaint with ROI-Background-Color*.



Figure 14: Left image shows benign items inpainted with Mean-Image-Color. The right image shows the same method with an additional filter applied to the benign regions.

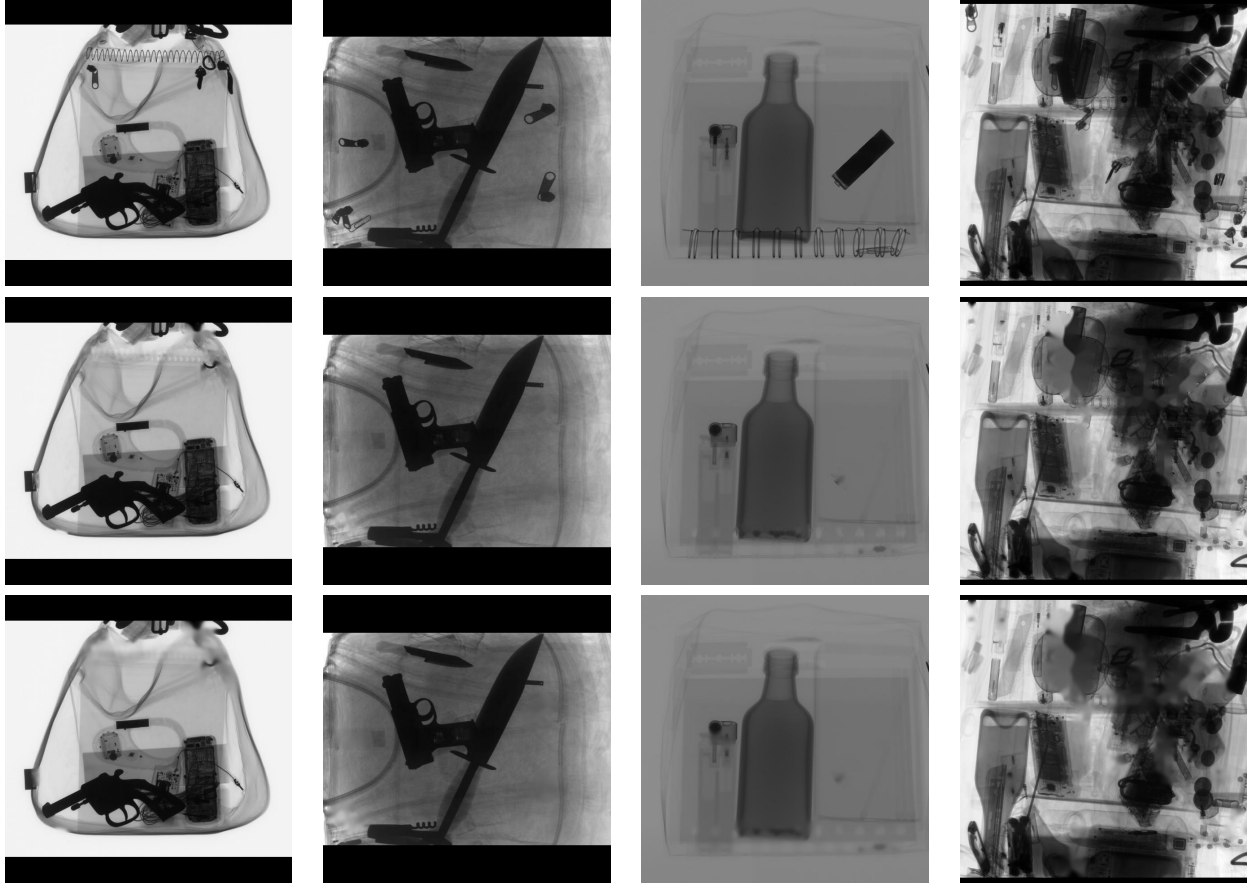


Figure 15: Coherent approaches - First row: original images; Second row: predicted benign object inpainted with Inpaint Coherent; Third row: additional filters applied to the inpainted images.

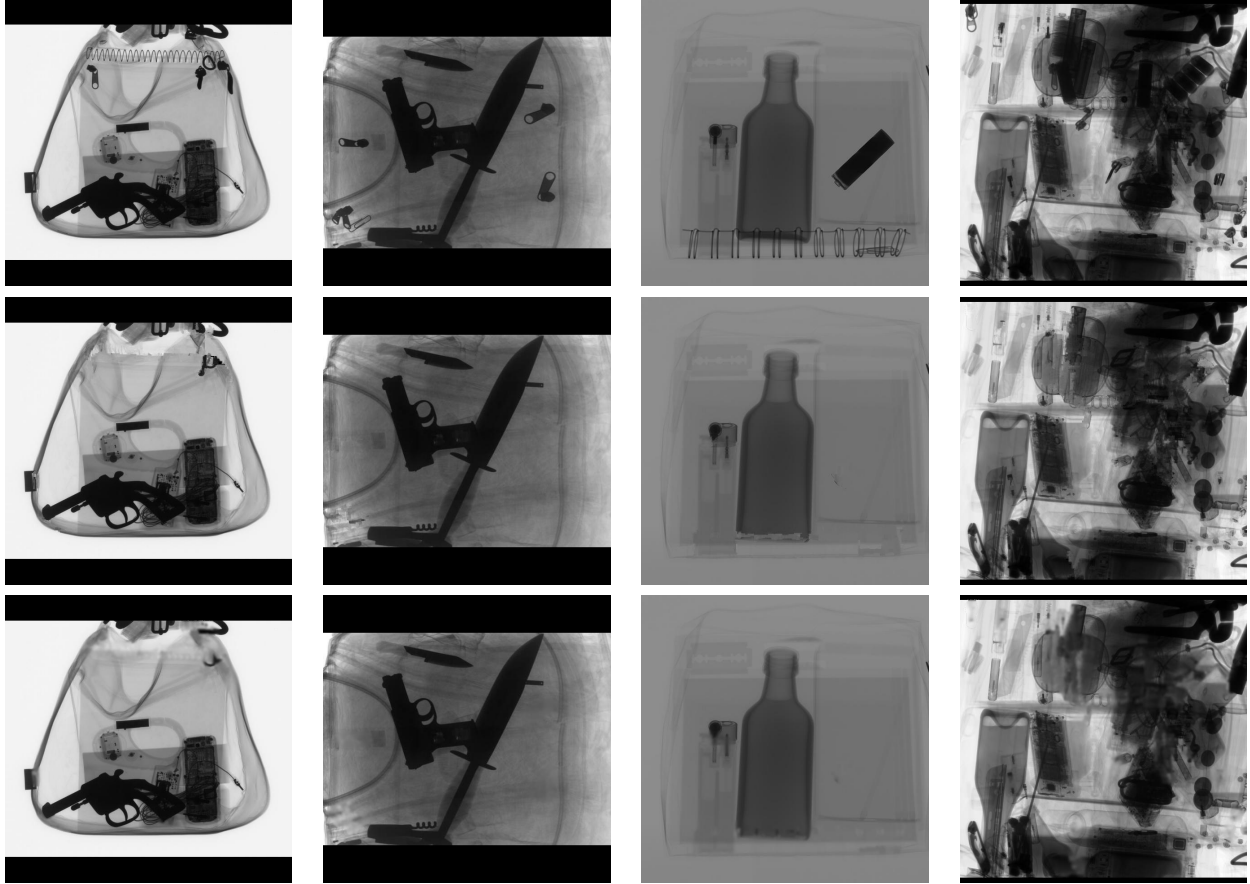


Figure 16: Exemplar approaches - First row: original images; Second row: predicted benign object inpainted with Inpaint Exemplar; Third row: additional filters applied to the inpainted images.

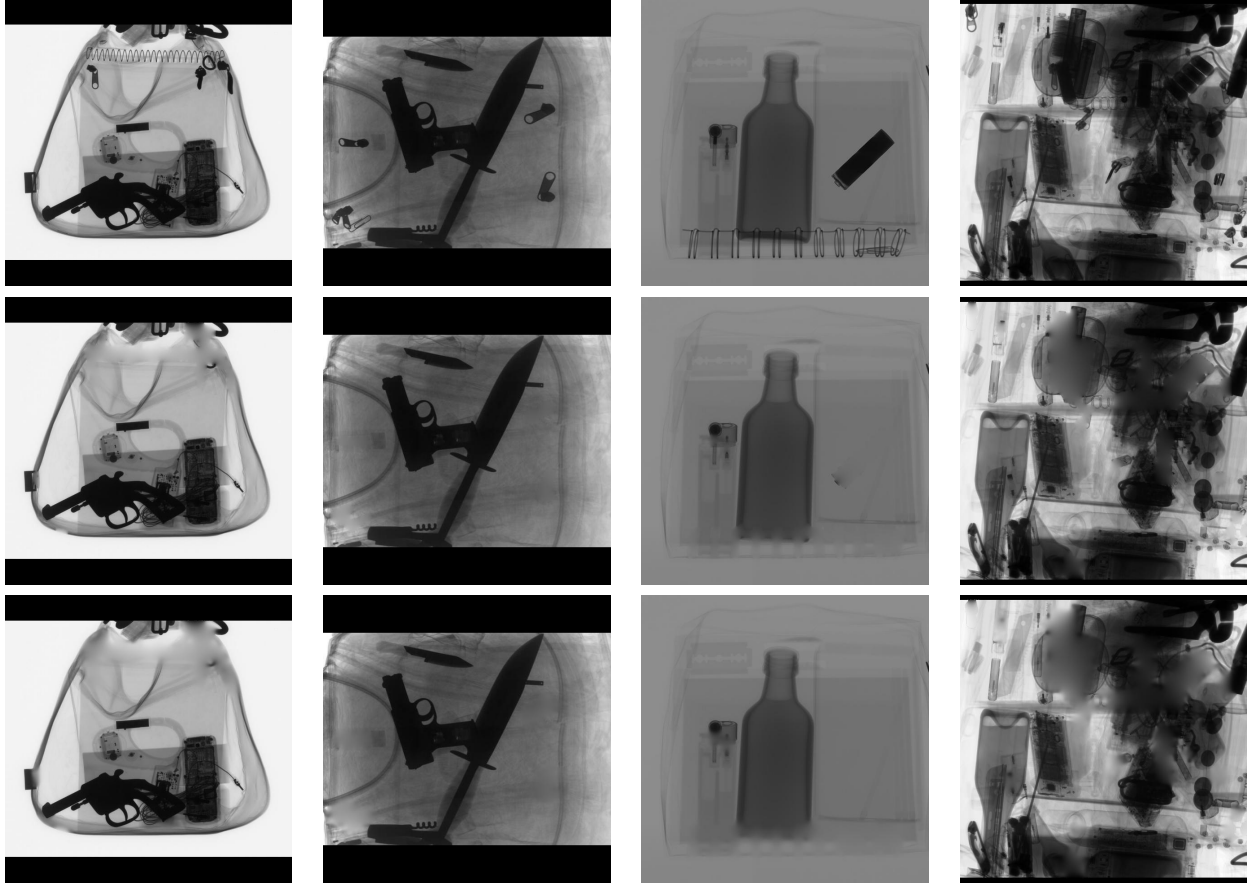


Figure 17: Regionfill approaches - First row: original images; Second row: predicted benign object inpainted with `Regionfill`; Third row: additional filters applied to the inpainted images.

5 Quantitative Evaluation

5.1 Results and Evaluation of the Object Detection Model

The trained object detection model is evaluated solely on the Test-set, a separate dataset to those used during the training and validation process. Example predictions on the Test-set are given in Figure 18.

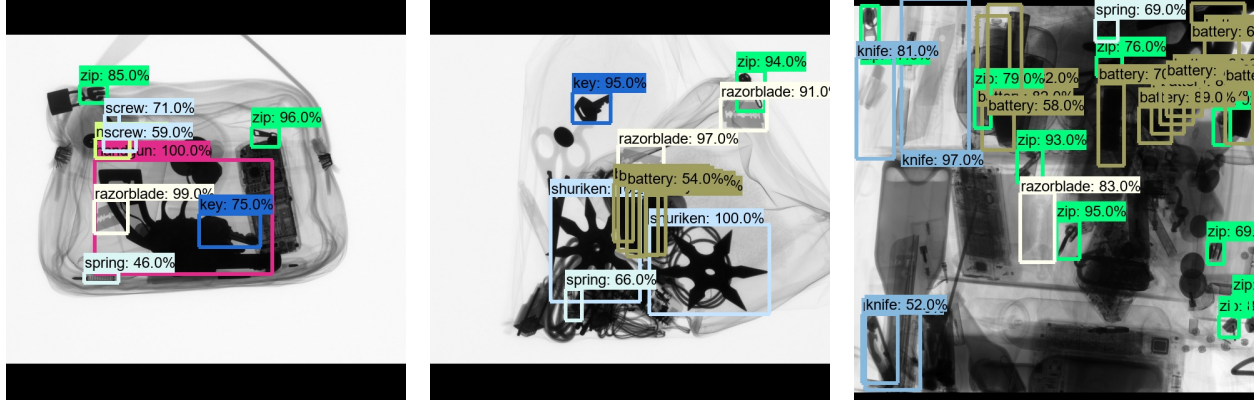


Figure 18: Example predictions by the model on the Test-set.

The Confusion Matrix given in Table 2 visualizes the model’s performance on the Test-set for all twelve categories. Noticeable in the Confusion Matrix are the many false-positive detections in background regions as well as multiple misses. The false-positive detections can be partly justified by multiple predictions for the same objects, for example, seen for the batteries in the two rightmost images in Figure 18. Too many predictions are especially the case for batteries and zips. Greedy nonmaximal suppression (NMS) could be used to eliminate overlapping bounding boxes for the same class labels. However, some objects in the data are very close together, and using NMS may result in an incorrect representation of correct predictions. The false-negative results for small objects (nail, screw, spring, staple, and paperclip) are likely due to a high scaling factor during training. Other misses could be due to many occlusions and severe clutter in the images. Some classes, such as the knife and screw, also contain objects of highly varying appearances, which could also contribute to wrong predictions. What emerges positively, however, is that classes are barely confused by the model. There are only a small number of objects that are mislabeled. The most problematic misclassification is the confusion of one knife as a battery. In this case, a threat object is classified as benign, leading to an incorrect removal of a threat item in the image enhancement process. Other derivations from the confusion matrix, such as the balanced accuracy (BA), are given in Table 3.

5.2 Quantitative Evaluation and Comparison of Distractor Removal Methods

Quantitative evaluation is applied to the enhanced images to determine, on the one hand, whether the distractor removal methods successfully reduced the overall visual clutter. On

Table 2: Confusion Matrix for the Multi-Class Detection on the Test-set; IoU threshold: 0.35; Score threshold: 0.45

		Predicted												
		Knife	Handgun	Shuriken	Razor Blade	Zip	Key	Nail	Screw	Spring	Staple	Battery	Paperclip	Miss
Ground Truth	Knife	22	0	0	0	0	0	0	0	0	0	1	0	14
	Handgun	0	66	0	0	0	0	0	0	0	0	0	0	0
	Shuriken	0	0	37	0	0	0	0	0	0	0	0	0	1
	Razor Blade	0	0	0	80	0	0	0	0	0	0	0	0	6
	Zip	0	0	0	0	239	0	0	0	0	0	0	0	71
	Key	0	0	0	1	0	35	0	0	0	0	0	0	13
	Nail	0	0	0	0	0	0	27	0	0	0	0	1	25
	Screw	0	0	0	0	0	0	0	49	0	0	2	0	75
	Spring	0	0	0	0	0	0	0	0	77	0	1	0	89
	Staple	0	0	0	0	0	0	0	0	0	8	0	0	56
	Battery	2	0	0	0	0	0	0	0	0	0	137	0	44
	Paperclip	0	0	0	0	0	0	0	0	0	0	0	22	47
	Background	14	9	1	3	32	5	9	14	4	0	56	11	0

the other hand, it should determine if the saliency decreased in the filtered regions while maintaining or even increasing salience in the areas containing threats.

5.2.1 Change in Visual Clutter

The overall visual clutter of the enhanced baggage scan is measured and compared to the visual clutter of the original image to determine whether the distractor removal methods successfully reduce visual clutter. To find out how the different distractor removal methods influence the visual clutter of the images, Quad-tree-clutter proposed by Jégou and Deblonde is applied [25, 64]. Quad-tree-clutter performs Quadtree Decomposition on a grayscale image. The image is subdivided into a quadtree based on the homogeneity of pixel values [35]. An example for such a decomposition is given in Figure 19. The global clutter value is then given by the number of cells in the resulting quadtree [64]. Quadtree Decomposition is performed by MATLAB’s function `qtdecomp(I, threshold)` [35]. The chosen threshold influences how often a block is subdivided into four. The smaller the threshold, the deeper the tree becomes, and thus the higher the resulting clutter value [64].

Table 4 shows how the different distractor removal methods affect visual clutter, where $\Delta Clutter$ denotes the change in clutter in reference to the original image ($\Delta Clutter = Clutter(I_{enhanced}) - Clutter(I_{original})$). The first column indicates the efficiency of the distractor removal methods in terms of visual clutter. It gives the ratio of images where clutter

Table 3: Derivations from the confusion matrix in Table 2.

	Knife	Handgun	Shuriken	Razor Blade	Zip	Key	Nail	Screw	Spring	Staple	Battery	Paperclip
Recall (TPR) (%)	59.5	100	97.4	93.0	77.1	71.4	50.9	38.9	46.1	12.5	74.9	31.9
Fall Out (FPR) (%)	1.2	0.7	0.1	0.3	2.9	0.4	0.7	1.1	0.3	0	4.9	0.9
Miss Rate (FNR) (%)	40.5	0	2.6	7.0	22.9	28.6	49.1	61.1	53.9	87.5	25.1	68.1
Precision (PPV) (%)	57.9	88.0	97.3	95.2	88.2	87.5	75.0	77.8	95.0	100	69.5	64.7
Accuracy (ACC) (%)	97.8	99.4	99.9	99.3	92.7	98.6	97.5	93.5	93.3	96.0	92.5	95.8
Balanced ACC (BA) (%)	79.1	99.7	98.6	96.4	87.1	85.5	75.1	68.9	72.9	56.3	85.0	65.5

is reduced to the total number of images.

All methods except the *Max-Background-Color* method reduce the clutter for at least 70% of the given images. The method *Filtered Regionfill* even achieves an efficiency of 98.5%, which means that this method reduces the visual clutter for 128 of the 130 total inputted images (only 130 images in total as images containing only threats were removed for this evaluation). *Filtered Regionfill* performs best in both efficiency and average clutter reduction. This method is closely followed by *Filtered Inpaint Coherent*, *Filtered Inpaint Mean-Image-Color* and *Filtered Inpaint Exemplar*. Noticeable is that filtering over the ROI after inpainting leads to a higher reduction in clutter within this experiment.

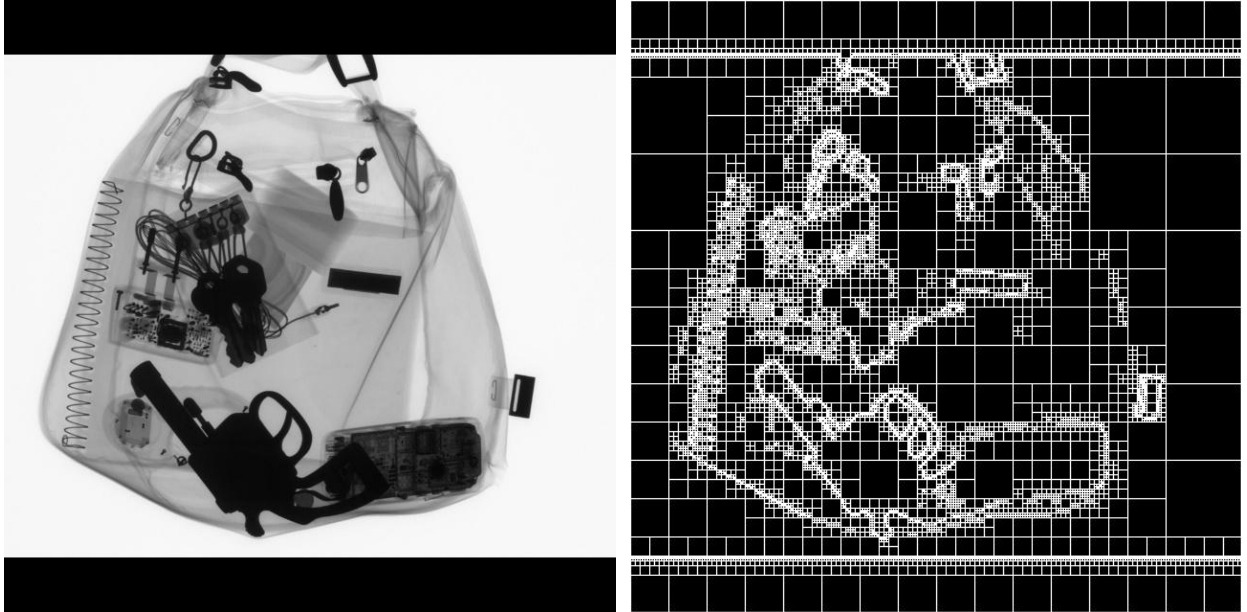


Figure 19: Quadtree Decomposition with threshold = 0.27.

Table 4: Clutter evaluated on Test-Set using predictions. 130 images in total (images containing only threats were removed for this evaluation).

	Efficiency	avg($\Delta Clutter$)	std($\Delta Clutter$)	min($\Delta Clutter$)	median($\Delta Clutter$)	max($\Delta Clutter$)
Inpaint Max-Background-Color	20.8%	889.5	1555.9	-1602	360	9387
Inpaint ROI-Background-Color	70.8%	-318.3	909.0	-3060	-153	2244
Inpaint Mean-Image-Color	73.1%	-345.3	768.3	-3036	-147	1239
Filtered Inpaint Mean-Image-Color	96.9%	-957.0	905.3	-3831	-669	273
Inpaint Coherent	90.8%	-756.1	824.4	-3342	-514.5	612
Filtered Inpaint Coherent	96.9%	-986.1	967.1	-3834	-676.5	468
Inpaint Exemplar	71.5%	-372.4	747.2	-3072	-159	1413
Filtered Inpaint Exemplar	96.2%	-942.8	940.9	-3816	-621	795
Regionfill	95.4%	-831.8	836.2	-3342	-570	336
Filtered Regionfill	98.5%	-1036.6	976.9	-3846	-732	252

5.2.2 Change in Saliency

The local changes in saliency are determined by measuring the saliency in the original image regions and comparing the values with the saliency in the same regions of the enhanced images. The local saliency is measured by applying the Itti-Koch-Niebur Saliency Model (IKN) to generate a saliency map for the whole image [23]. The middle column in Figure 20 shows two different saliency maps produced by IKN. The used MATLAB implementation is provided by the *Saliency Model Implementation Library for Experimental Research* (SMILER) [69].

Figures 21 and 22 show, for each method, the average decrease in saliency in regions of former distractors and the increase in saliency in threat regions, each as a percentage of the original saliency. *Filtered Regionfill* achieves the highest reduction of benign-saliency with an average of 31.9%. All methods achieve an average increase of threat-saliency of at least 1%. It is important to note, however, that the weighted average is susceptible to outliers.

Table 5 shows further results that indicate how the different Distractor Removal methods affect visual saliency in regions containing benign objects (first three columns) and regions containing threats (last three columns).

$\Delta SaliencyMap$ denotes the saliency change in an enhanced image in reference to the original one where $\Delta SaliencyMap = SaliencyMap_{enhanced} - SaliencyMap_{original}$.

The last column of Figure 20 visualizes the $\Delta SaliencyMap$ between the $SaliencyMap_{original}$ and $SaliencyMap_{enhanced}$ shown in the middle column.

$\Delta B_{\text{SaliencyMap}}$ denotes the changes in salience within the benign region of an enhanced image. The smaller the values, the better as salience should be reduced in benign regions.

$$\Delta B_{\text{SaliencyMap}} = \text{SaliencyMap}_{\text{enhanced}}(\text{BenignROI}) - \text{SaliencyMap}_{\text{original}}(\text{BenignROI})$$

$\Delta T_{\text{SaliencyMap}}$ denotes the changes in salience within the threat region of an enhanced image. The greater the values, the better as salience should be increased in threat regions.

$$\Delta T_{\text{SaliencyMap}} = \text{SaliencyMap}_{\text{enhanced}}(\text{ThreatROI}) - \text{SaliencyMap}_{\text{original}}(\text{ThreatROI})$$

$\text{weightedAvg}(\Delta \text{RegionSaliencyMap})$ denotes the average change in salience over the given image region. The average change in salience within a single image is calculated in MATLAB by $\text{mean2}(\Delta \text{RegionSaliencyMap})$. $\text{mean2}(\mathbf{A})$ computes the mean of the given array \mathbf{A} [35]. To summarize the mean values overall images, the weighted average must be used because the region of interest changes between images. Therefore, the area of the region may change. Each mean value is weighted by the region’s area (size of array \mathbf{A}). The summarized value is desired to be negative for benign regions, while it should be positive for threat regions.

image-wise efficiency($\Delta \text{RegionSaliencyMap}$) gives the ratio of images where salience in benign/threat regions is successfully reduced/increased to the total input images (130). Whether an image is classified as successful is decided based on the $\text{mean2}(\Delta \text{RegionSaliencyMap})$. $\text{Mean} < 0$ denotes that the salience in the region is reduced, while $\text{mean} \geq 0$ denotes that the salience is increased.

pixel-wise efficiency($\Delta \text{RegionSaliencyMap}$) gives the percent of pixels where salience in benign/threat regions is successfully reduced/increased. As the number of considered pixels changes for each image, the weighted average has to be used to summarize the values of overall images. The weights are the area of the considered ROI’s.

The chosen distractor removal methods aim to decrease salience in the enhanced regions while maintaining or even increasing salience regions containing threats. The only method that failed to fulfill this requirement on the Test-set using predictions is the *Inpaint Max-Background-Color* method. Instead of decreasing salience in the enhanced regions, it increased it on average. The methods that worked best on the evaluated set are *Regionfill* and *Filtered Regionfill*. These results are consistent with the clutter measurements.

5.3 Results of the Feedback Loop

Since distractor removal discards basic features such as edges and contrast from the processed regions, it could influence CNNs. Whether the detection model is affected can be determined by feeding the enhanced images as input to the model. The results then can be compared to the performance obtained with the original images. This evaluation is done on a subset of the test set. All images containing only a single item were removed, resulting in a subset containing 101 images.

As the primary goal of the distractor removal methods is to filter out distractive items, inpainted benign items should not be detectable anymore. Therefore, one crucial question is whether the detection model correctly rejects the removed benign items. The percentage of correctly rejected benign items is measured by first identifying and counting all benign items removed in the distractor removal process that the model still detects. The percentage is

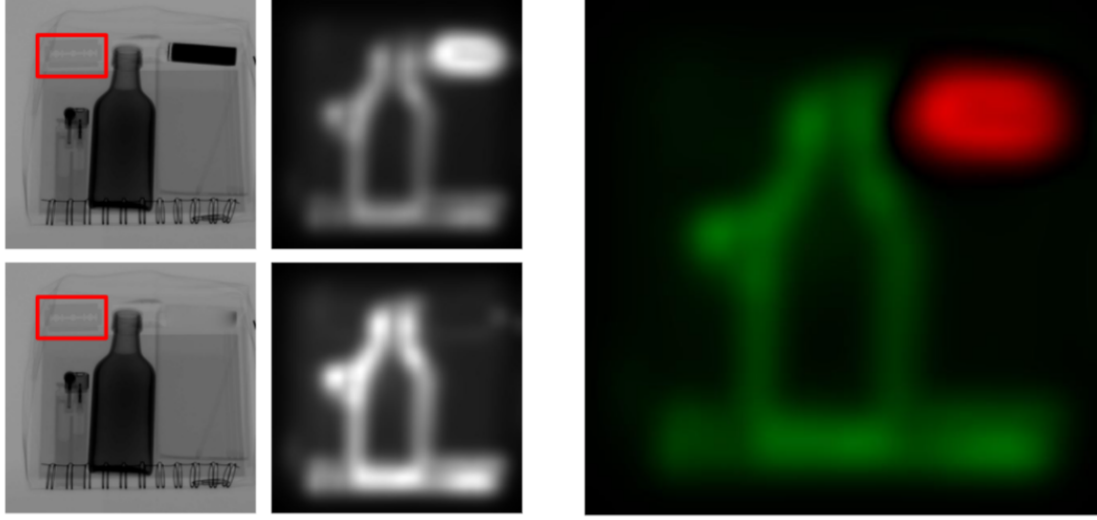


Figure 20: Visualization of Saliency Maps and $\Delta SaliencyMap$.

The original image and its enhanced version are displayed on the left, containing both a marked threat object. The images in the middle show the salience maps, and the right image visualizes $\Delta SaliencyMap$ of the two maps. Red denotes a reduction in salience, and green an increase.

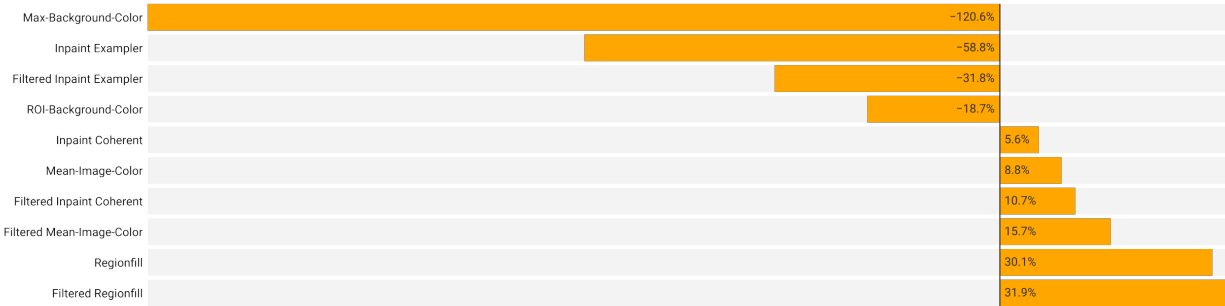


Figure 21: Decrease of saliency in regions of former distractors as a percentage of the original saliency. For each image, the percentage is calculated pixel-wise and then averaged. The final value is the weighted average overall images based on the number of pixels in the regions of interest.

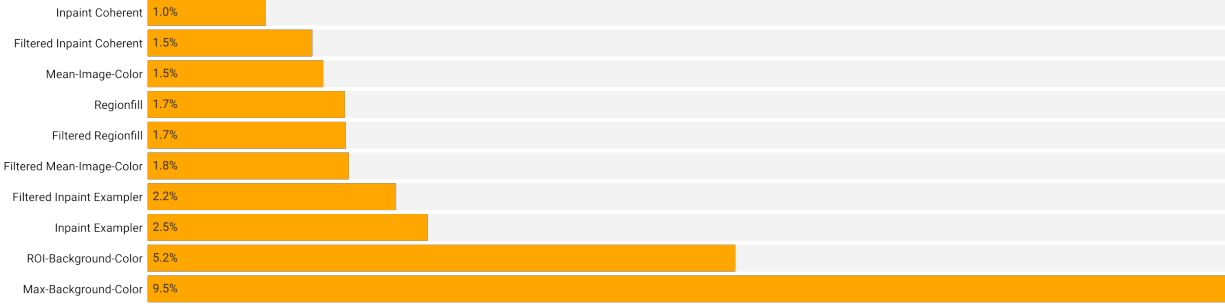


Figure 22: Increase of saliency in threat regions as a percentage of the original saliency. For each image, the percentage is calculated pixel-wise and then averaged. The final value is the weighted average overall images based on the number of pixels in the regions of interest.

then given by $1 - (\text{detected_distractors} / \text{total_distractors})$. The total number of distractors removed from the dataset is 583. As shown in Figure 24, *Filtered Regionfill* performs best with 88.87% correct rejections. This result, however, also means that at least 11.13% of the removed benign items are still detectable by the model. This can be partly explained by benign objects overlapping threats. Such overlaps cannot be processed, otherwise, there is a risk that the threat will become unrecognizable. Furthermore, semantic segmentation may fail when benign and threats are close together. Regarding the correct-rejection-rate, the Max-Background-Color, Mean-Image-Color, and ROI-Background-Color methods do not perform as well as the others, presumably because the shapes of the filtered objects are still very prominent even after removing the distractors.

All distractor removal methods lead to an increase in true-positive detections of benign items as plotted in Figure 23. 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%. Examples of additional benign item detections are given in Figure 25.

Table 5: Saliency evaluated on Test-Set using predictions. 130 images in total (images containing only threats were removed for this evaluation). Value in the saliency map ranges from 0 to 255.

	$\text{weightedAvg}(\Delta B_{\text{SaliencyMap}})$	$\text{image-wise efficiency}(\Delta B_{\text{SaliencyMap}})$	$\text{pixel-wise efficiency}(\Delta B_{\text{SaliencyMap}})$	$\text{weightedAvg}(\Delta T_{\text{SaliencyMap}})$	$\text{image-wise efficiency}(\Delta T_{\text{SaliencyMap}})$	$\text{pixel-wise efficiency}(\Delta T_{\text{SaliencyMap}})$
Inpaint Max-Background-Color	+16.1	31.5%	41.4%	-2.1	51.5%	39.9%
Inpaint ROI-Background-Color	-20.8	83.1%	69.4%	+1.4	56.9%	51.7%
Inpaint Mean-Image-Color	-24.8	83.8%	74.6%	+0.57	56.2%	55.5%
Filtered Inpaint Mean-Image-Color	-28.7	91.5%	77.8%	+0.63	54.6%	53.6%
Inpaint Coherent	-24.4	89.2%	65.2%	+0.3	54.6%	49.5%
Filtered Inpaint Coherent	-26.6	90.8%	68.3%	+0.5	54.6%	50.5%
Inpaint Exemplar	-14.0	73.1%	54.9%	+1.1	60.8%	54.2%
Filtered Inpaint Exemplar	-20.9	82.3%	62.9%	+0.9	59.2%	53.8%
Regionfill	-32.8	96.9%	82.3%	+0.8	61.5%	53.0%
Filtered Regionfill	-34.0	96.2%	83.8%	+0.9	53.8%	52.0%

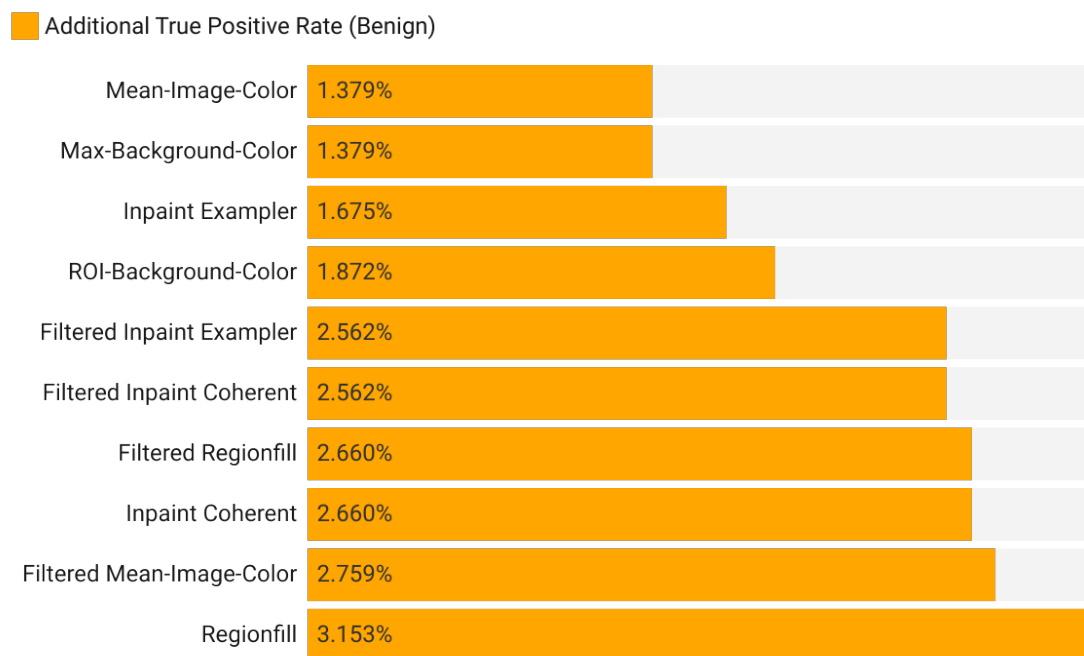


Figure 23: Percentage of additional benign items relative to the total number (1015) detected after applying distractor removal methods.

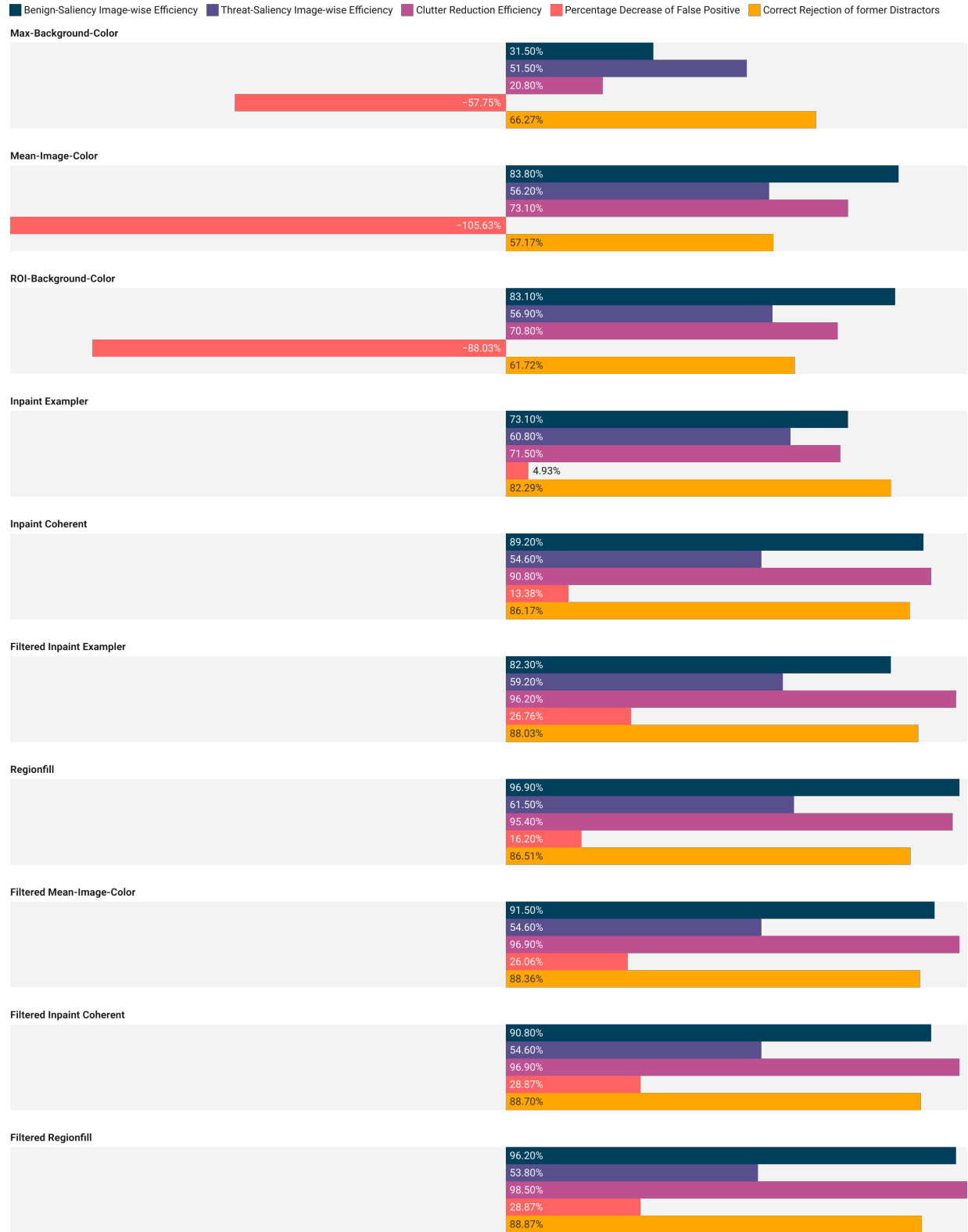


Figure 24: Final ranking of the distractor removal methods from worst at the top to best at the bottom.

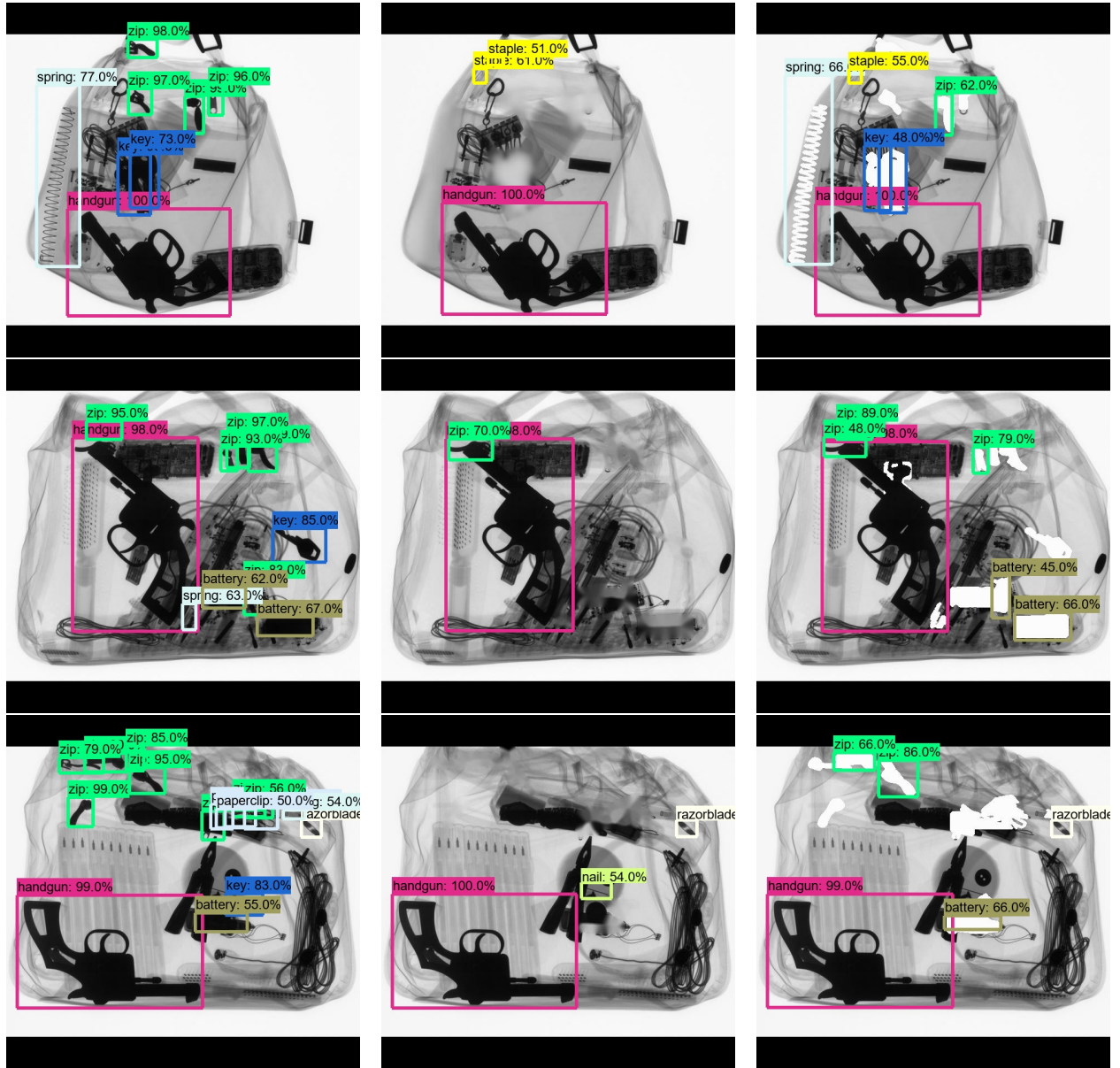


Figure 25: First column: predictions on original images; Second column: prediction on enhanced images where new correct predictions are done; Third column: prediction on enhanced images where *Inpaint Max-Background-Color* was used (only method that failed to reduce saliency and clutter).

6 Conclusion and Future Work

In this work, the concept of detecting and removing benign items from baggage scans is presented and experimentally evaluated. Supplemental detection of benign objects provides additional information that can be used to assist baggage screening and offers several advantages over sole threat detection. Eliminating distractions during baggage screening is potentially valuable for focusing attention on regions containing threats or shortening the visual search task by reducing visual clutter and altering the saliency of specific image regions.

A strong argument for benign detection is that it is more feasible without human operators than threat detection in a security context. In baggage screening, prohibited items must not be overlooked. As far as possible, all threats have to be detected by automatic threat detection or by human operators without misses. However, the same does not apply to benign items. Not all types of benign objects need to be identified, but the detected ones have to be accurate. Furthermore, contrary to threats, benign items are not usually deliberately concealed, making them easier to detect correctly. In addition to the proposed application, the information gained from benign detection could be used in various ways. One possibility is to apply benign detection as a diagnostic aid, for instance, by highlighting benign items that often need to be checked manually by human operators, such as laptops or other electronic devices.

The experiments demonstrate that removing distracting items positively influences the scans' saliency. In up to 96.9% of the tested images, the saliency in regions of former distractors is reduced by up to 31.94%. On the other hand, the saliency in threat regions is increased by about 2% in up to 61.5% of the images. Furthermore, all distractor removal methods except Max-Background-Color successfully reduced the clutter for at least 70% of the given images. The best method achieves an efficiency of 98.5% on the test set. These 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 another experiment showing that removing distractors positively impacts our detection model. After applying distractor removal methods, the rate of true-positive items increases from 58.42% to a maximum of 61.58%, indicating that the model can identify additional items after distractor removal. Additionally, false-positive detections were decreased, at most by 28.87%. Moreover, the model correctly rejects removed distractors in up to 88.87% of the cases. Some of the distractors are still detectable, as they overlay threat items and therefore are not possible to exclude entirely. From this result arises further possible research directions for future work. For example, whether benign detection and distractor removal could be used to support other automated systems by providing refined inputs.

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, as this gives the impression that the enhanced regions are empty. This disadvantage is omitted as soon as more information about the bag is available, such as when working with computer tomographs that provide 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 cause an undesirable effect. Specifically, removing nails from a nail bomb may make it harder to recognize the bomb.

Finally, the experiments suggest a potential for benign item detection and distractor removal. However, further studies about the practical application of benign detection are required. Additionally, future research into distractor removal and, in particular, its potential in 3D would be interesting.

References

- [1] Samet Akcay and Toby Breckon. Towards automatic threat detection: A survey of advances of deep learning within x-ray security imaging. *arXiv preprint arXiv:2001.01293*, 2020.
- [2] Ilhan AYDIN, Mehmet KARAKOSE, and AKIN Erhan. A new approach for baggage inspection by using deep convolutional neural networks. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*, pages 1–6. IEEE, 2018.
- [3] Muhammet Baştan. Multi-view object detection in dual-energy x-ray images. *Machine Vision and Applications*, 26(7):1045–1060, 2015.
- [4] Emil Benedykciuk, Marcin Denkowski, and Krzysztof Dmitruk. Material classification in x-ray images based on multi-scale cnn. *Signal, Image and Video Processing*, pages 1–9, 2021.
- [5] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, and Coloma Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424, 2000.
- [6] Adam T Biggs, Stephen H Adamo, and Stephen R Mitroff. Rare, but obviously there: Effects of target frequency and salience on visual search accuracy. *Acta psychologica*, 152:158–165, 2014.
- [7] Folkmar Bornemann and Tom März. Fast image inpainting based on coherence transport. *Journal of Mathematical Imaging and Vision*, 28(3):259–278, 2007.
- [8] Chen Chen, Xianzhi Du, Le Hou, Jaeyoun Kim, Pengchong Jin, Jing Li, Yeqing Li, Abdullah Rashwan, and Hongkun Yu. Tensorflow official model garden, 2020.
- [9] Marvin M Chun. Contextual cueing of visual attention. *Trends in cognitive sciences*, 4(5):170–178, 2000.
- [10] Christine Connolly. X-ray systems for security and industrial inspection. *Sensor Review*, 28(3):194–198, 2008.
- [11] Antonio Criminisi, Patrick Pérez, and Kentaro Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing*, 13(9):1200–1212, 2004.
- [12] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR’05)*, volume 1, pages 886–893. IEEE, 2005.
- [13] Nick Donnelly, Alex Muhl-Richardson, Hayward J Godwin, and Kyle R Cave. Using eye movements to understand how security screeners search for threats in x-ray baggage. *Vision*, 3(2):24, 2019.

- [14] Thorsten Franzel, Uwe Schmidt, and Stefan Roth. Object detection in multi-view x-ray images. In *Joint DAGM (German Association for Pattern Recognition) and OAGM Symposium*, pages 144–154. Springer, 2012.
- [15] Ohad Fried, Eli Shechtman, Dan B Goldman, and Adam Finkelstein. Finding distractors in images. In *Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [16] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.
- [17] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.
- [18] Lewis D Griffin, Matthew Caldwell, and Jerone T A Andrews. Compass-xp dataset, 2019. data retrieved from, <https://doi.org/10.5281/zenodo.2654887>.
- [19] Antonio Gulli and Sujit Pal. *Deep learning with Keras*. Packt Publishing Ltd, 2017.
- [20] PA Hancock and SG Hart. Defeating terrorism: What can human factors/ergonomics offer? *Ergonomics in design*, 10(1):6–16, 2002.
- [21] Douglas H Harris. How to really improve airport security. *Ergonomics in Design*, 10(1):17–22, 2002.
- [22] Nicole Hättenschwiler, Yanik Sterchi, Marcia Mendes, and Adrian 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.
- [23] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 20(11):1254–1259, 1998.
- [24] Deepak Kumar Jain et al. An evaluation of deep learning based object detection strategies for threat object detection in baggage security imagery. *Pattern Recognition Letters*, 120:112–119, 2019.
- [25] Laurent Jégou and Jean-Philippe Deblonde. Vers une visualisation de la complexité de l’image cartographique. *Cybergeo: European Journal of Geography*, 2012.
- [26] LS Johns, A Shaw, and A Fainberg. Technology against terrorism: The federal effort. *The OTA Report*, pages 78–80, 1991.
- [27] Ronald T Kneusel and Michael C Mozer. Improving human-machine cooperative visual search with soft highlighting. *ACM Transactions on Applied Perception (TAP)*, 15(1):1–21, 2017.
- [28] Saskia M Koller, Colin G Drury, and Adrian Schwaninger. Change of search time and non-search time in x-ray baggage screening due to training. *Ergonomics*, 52(6):644–656, 2009.

- [29] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.
- [30] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755. Springer, 2014.
- [31] Xi Liu, Alastair Gale, and Tao Song. Detection of terrorist threats in air passenger luggage: Expertise development. In *2007 41st Annual IEEE International Carnahan Conference on Security Technology*, pages 301–306. IEEE, 2007.
- [32] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.
- [33] Qiang Lu. *The utility of X-ray dual-energy transmission and scatter technologies for illicit material detection*. Virginia Polytechnic Institute and State University, 1999.
- [34] Victor A Mateescu and Ivan V Bajic. Visual attention retargeting. *IEEE MultiMedia*, 23(1):82–91, 2015.
- [35] MATLAB. *version 9.8.0.1380330 (R2020a)*. The MathWorks Inc., Natick, Massachusetts, 2020.
- [36] Nelson Max. Optical models for direct volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 1(2):99–108, 1995.
- [37] Jason S McCarley, Arthur F Kramer, Christopher D Wickens, Eric D Vidoni, and Walter R Boot. Visual skills in airport-security screening. *Psychological science*, 15(5):302–306, 2004.
- [38] Roey Mechrez, Eli Shechtman, and Lihi Zelnik-Manor. Saliency driven image manipulation. *Machine Vision and Applications*, 30(2):189–202, 2019.
- [39] Domingo Mery. Computer vision for x-ray testing. *Switzerland: Springer International Publishing*, pages 1–5, 2015.
- [40] Domingo Mery, Vladimir Rizzo, Uwe Zscherpel, German Mondragón, Iván Lillo, Irene Zuccar, Hans Lobel, and Miguel Carrasco. Gdxdxray: The database of x-ray images for nondestructive testing. *Journal of Nondestructive Evaluation*, 34(4):42, 2015.
- [41] Domingo Mery, Daniel Saavedra, and Mukesh Prasad. X-ray baggage inspection with computer vision: A survey. *IEEE Access*, 8:145620–145633, 2020.
- [42] Domingo Mery, Erick Svec, Marco Arias, Vladimir Rizzo, Jose M Saavedra, and Sandipan Banerjee. Modern computer vision techniques for x-ray testing in baggage inspection. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 47(4):682–692, 2017.
- [43] Caijing Miao, Lingxi Xie, Fang Wan, chi Su, Hongye Liu, jianbin Jiao, and Qixiang Ye. Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images. In *CVPR*, 2019.

- [44] Stefan Michel, Saskia Koller, Markus Ruh, and Adrian Schwaninger. The effect of image enhancement functions on x-ray detection performance. In *Proceedings of the 4th International Aviation Security Technology Symposium*, 11 2006.
- [45] Stefan Michel and Adrian Schwaninger. Human-machine interaction in x-ray screening. In *Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on*, pages 13–19. IEEE, 2009.
- [46] Keiron O’Shea and Ryan Nash. An introduction to convolutional neural networks. *arXiv preprint arXiv:1511.08458*, 2015.
- [47] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, 9(1):62–66, 1979.
- [48] Niall O’Mahony, Sean Campbell, Anderson Carvalho, Suman Harapanahalli, Gustavo Velasco Hernandez, Lenka Krpalkova, Daniel Riordan, and Joseph Walsh. Deep learning vs. traditional computer vision. In *Science and Information Conference*, pages 128–144. Springer, 2019.
- [49] Raman B Paranjape. Fundamental enhancement techniques. *Handbook of medical imaging*, pages 3–18, 2000.
- [50] Nathalie Pattyn, Xavier Neyt, David Henderickx, and Eric Soetens. Psychophysiological investigation of vigilance decrement: boredom or cognitive fatigue? *Physiology & behavior*, 93(1-2):369–378, 2008.
- [51] Vladimir Rizzo, Sebastian Flores, and Domingo Mery. Threat objects detection in x-ray images using an active vision approach. *Journal of Nondestructive Evaluation*, 36(3):1–13, 2017.
- [52] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.
- [53] Adrian Schwaninger. X-ray imagery: Enhancing the value of the pixels. *Aviation Security International*, pages 16–21, 10 2005.
- [54] Adrian Schwaninger and Sarah Merks. Single-view, multi-view and 3d imaging for baggage screening: What should be considered for effective training? *Aviation Security International*, 2019.
- [55] Adrian Schwaninger, Stefan Michel, and Anton Bolting. Towards a model for estimating image difficulty in x-ray screening. In *Proceedings 39th Annual 2005 International Carnahan Conference on Security Technology*, pages 185–188. IEEE, 2005.
- [56] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep convolutional neural networks for computer-aided detection: Cnn architectures, dataset characteristics and transfer learning. *IEEE transactions on medical imaging*, 35(5):1285–1298, 2016.

- [57] Jacek Skorupski and Piotr Uchroński. A human being as a part of the security control system at the airport. *Procedia Engineering*, 134:291–300, 2016.
- [58] Malcolm Sperrin and John Winder. *Scientific Basis of the Royal College of Radiologists Fellowship*. IOP Publishing, 2014.
- [59] Yanik Sterchi and Adrian Schwaninger. A first simulation on optimizing eds for cabin baggage screening regarding throughput. In *2015 International Carnahan conference on security technology (ICCST)*, pages 55–60. IEEE, 2015.
- [60] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.
- [61] Mingxing Tan, Ruoming Pang, and Quoc V Le. Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10781–10790, 2020.
- [62] Shijian Tang and Ye Yuan. Object detection based on convolutional neural network. In *International Conference-IEEE-2016*, 2015.
- [63] Renful Premier Technologies. X-ray screener: X-ray imaging. Website. Online erhältlich unter <http://www.x-rayscreener.co.uk/?xray=x-ray-imaging> last accessed on 22 January 2018.
- [64] Guillaume Touya, Blandine Decherf, Mayeul Lalanne, and Marion Dumont. Comparing image-based methods for assessing visual clutter in generalized maps. *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2:227–233, 2015.
- [65] Diana Turcsany, Andre Mouton, and Toby 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.
- [66] Tzutalin. Labelimg. Free Software: MIT License, 2015.
- [67] Joel S Warm, Raja Parasuraman, and Gerald Matthews. Vigilance requires hard mental work and is stressful. *Human factors*, 50(3):433–441, 2008.
- [68] K Wells and DA Bradley. A review of x-ray explosives detection techniques for checked baggage. *Applied Radiation and Isotopes*, 70(8):1729–1746, 2012.
- [69] Calden Wloka, Toni Kunić, Iuliia Kotseruba, Ramin Fahimi, Nicholas Frosst, Neil DB Bruce, and John K Tsotsos. Smiler: Saliency model implementation library for experimental research. *arXiv preprint arXiv:1812.08848*, 2018.
- [70] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232, 2019.