Automated Threat Detection In X-Ray Imagery For Advanced Security Applications
---- MM811 Literature Review

Rajat Bothra Jain, Amit Upreti
Nov 22nd, 2021

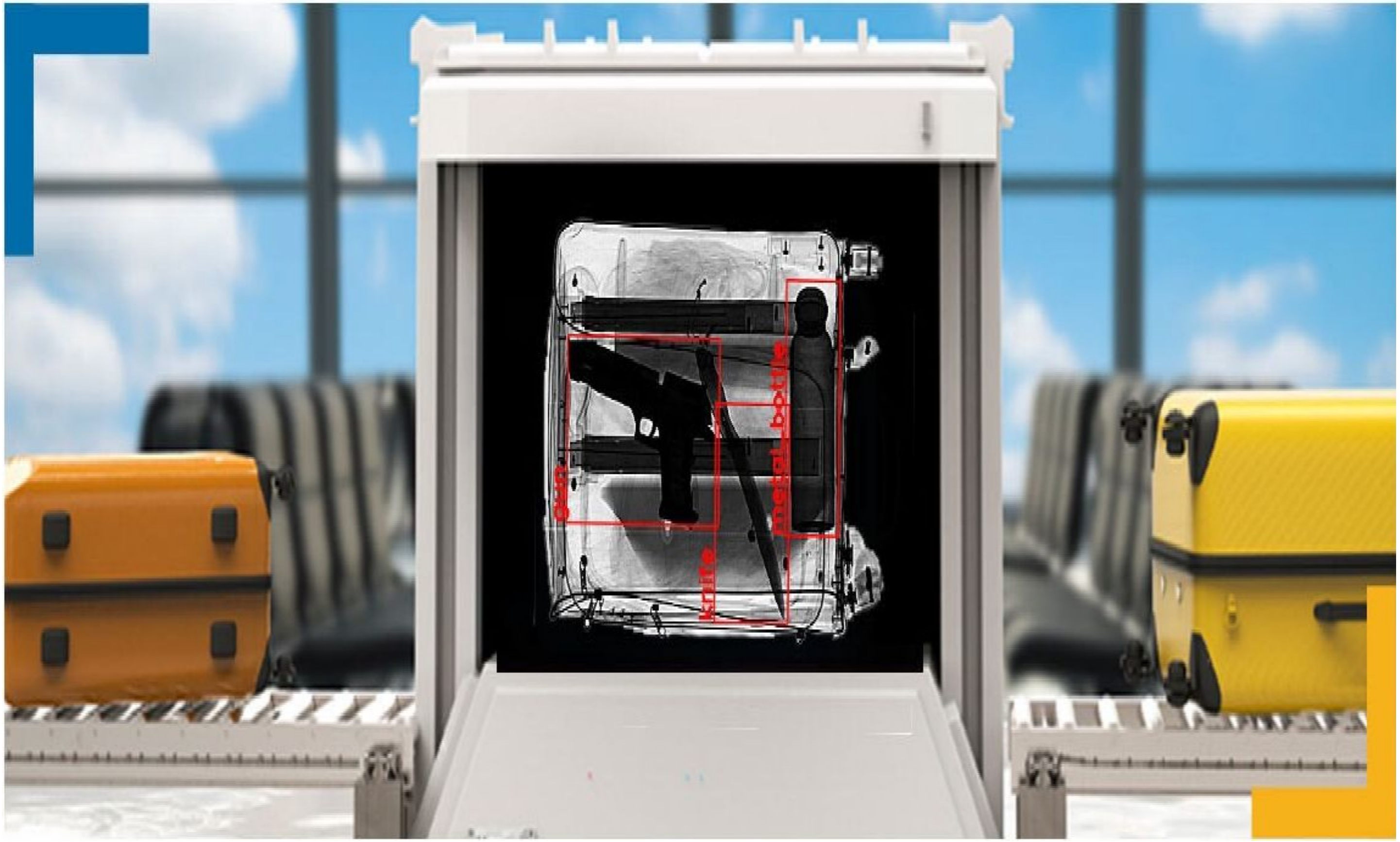
UNIVERSITY OF ALBERTA

Master Program
Department of Computing Science

Contents

- Introduction
- Literature Review

Disrupting Conventional Baggage Scanning with AI



https://medium.com/baggageai/disrupting-conventional-baggage-scanning-with-ai-b928ada2f721

Introduction

Where Baggage Scan is used ?

X-Ray machines are used to scan baggage to look for hazardous objects such as guns, knives, razor blades in areas of high security, like airports, railway station etc.

Why automatic threat detection in Baggage Scan is important?

- Time & Effort Reduction
- Avoid Human Errors

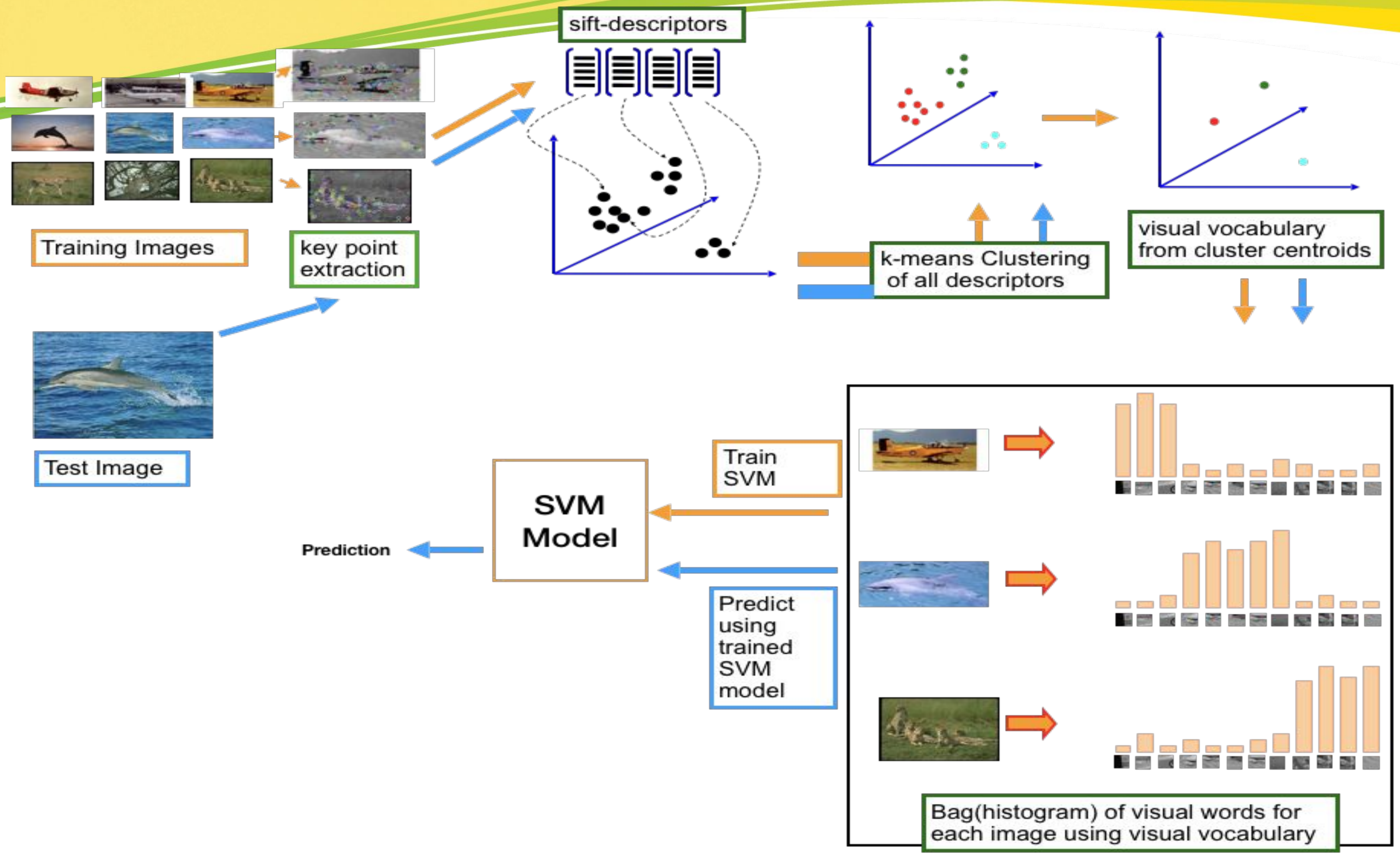
Computer Vision Generic object detection and classification problem

Object Detection & Classification

Neural Based Approach : Typically based on Convolutional Neural Network

Non-Neural Based Approach : First define the features using HOG, SIFT etc. than do classification by technique such as SVM

Visual Words on Baggage X-Ray Images [General Framework]



Dataset : For each 3 types of Images (low-energy, high energy, color image) we have 4 views (Top,Side and Two Views From Some Angle) hence 12 images for each baggage.

Point Detectors : DoG, Hessian-Laplace, Harris, FAST and STAR

Descriptors: SIFT , SURF and BRIEF

Low Accuracy with Bag Of Words :

X-ray images data is more challenging than visual image data because they lack textures , which is essential for Bag of words.

a) they are transparent, pixel values represent the attenuation by multiple objects

b) they may be very cluttered, which dramatically increases the no of meaningless interest points.

c) Noisy due to low energy X-ray imaging

Changes To Improve Accuracy: Trying To Find Better Feature Detector

- Using 12 images information to decide the threat.
- SIFT Performed best among all descriptors
- Color segmentation and filtering out the background keypoints.

- Union of point detectors and concatenation of BoWs (color filtered keypoints + SIFT) provides best results.

Feature Detection & Descriptor :

Pre-processing : Foreground segmentation by truncating greyscale pixel intensities above/below a threshold value, prior to applying the detection algorithm such as SURF. **[Cluttered Image: Noisy Points]**

Low Threshold : Saliency measure for interest point identification resulting in a higher number of generated features. **[Lack Texture: Less Points]**

Visual codebook generation:

X-ray images into two classes: positive which represent the target object and negative which represent background i.e. all ‘target-free’ images.Primed BoW representation simplifies the task of the classifier. For visual words (k) we will have 2k size codebook.

Fig. 3. Training data examples. Positive instances (left) showing firearms of various sizes, shapes, orientations etc. Negative instances (right) showing a variety of clutter items.

Results Comparison:

This work results (TPR = 99.07%) compared to the optimal results (TPR = 70%) presented in Bastan.et al work. Both approaches differs in particulars below but Basten.et al choice of components is based on prior experimentation thus rule out difference in SURF descriptors and the RBF kernel

	This study	Bastan *et al.* [6]
Pre-processing	foreground segmentation	foreground segmentation
Interest point detection	SURF	DoG + Harris
Descriptor	SURF	SIFT
Codebook generation	class-specific online *k*-means clustering	traditional *k*-means
Size of codebook (number of clusters)	1200	200
Vector quantisation	Hard assignment	Soft assignment
Classifier	SVM	SVM
Kernel	Gaussian RBF	Histogram intersection
Experimentation	3-fold cross validation	Separate training and test sets
Scanner	Dual-view, single energy X-ray (2 images per item)	Quad-view dual-energy X-ray (12 images per item)
Data	850 pos; 10000 neg	208 training (52 pos; 156 neg) 764 test (40 pos; 724 neg)

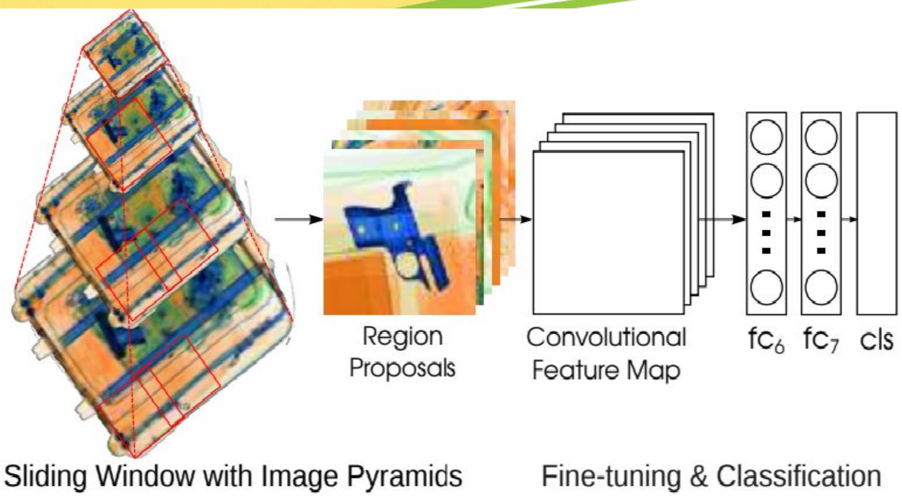
Table 2. Comparison of proposed approach and method of Bastan *et al.* [6]

Sliding Window Based Approaches :Generate region proposals by sliding a fixed size window through the image. With a combination with image pyramids for varying sizes and scales fed into a CNN for the classification step.

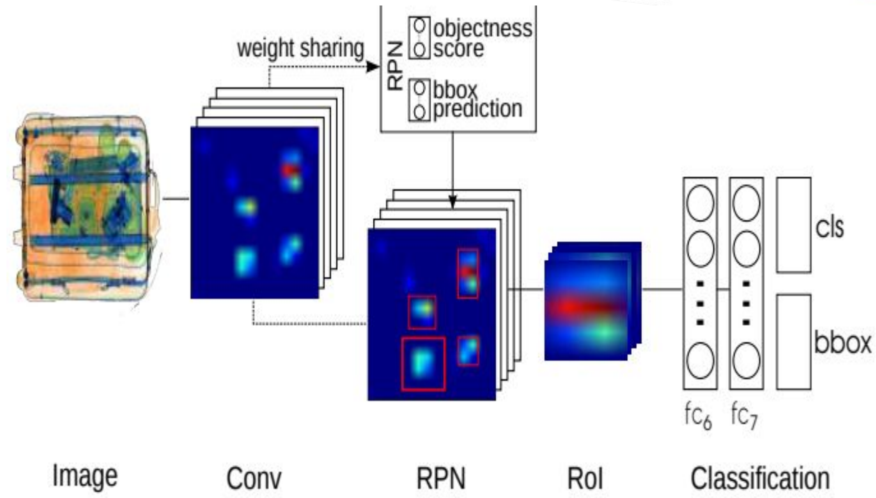
Faster RCNN : Two subnetworks, containing a unique region proposal network (RPN) and Fast RCNN network together. Convolutional layers in RPN are shared with Fast RCNN network, which makes region proposal step cost efficient compared to using an external region proposal algorithm.Features of object candidates are generated by sliding a window over the feature map of the last convolutional layer. Using these features, the last two fully connected layers (bounding box regression and classification) then generate region proposals. RoI pooling layer resizes the proposals to have fixed sized width and length. fc layers then create feature vector to be used by bounding box regression and softmax layers

R-FCN : Removed out the main limitation of Faster RCNN that each region proposal within RoI pooling layer is computed hundreds of times.They propose a new approach removing fully connected layers after RoI pooling, and employing a new approach called "position sensitive score map" which handles translation variance issue in detection task. Since no fully connected subnetwork is used, the proposed model shares weights within almost entire network. This leads to much faster convergence both in training and test stages, while achieving similar results to Fast RCNN.

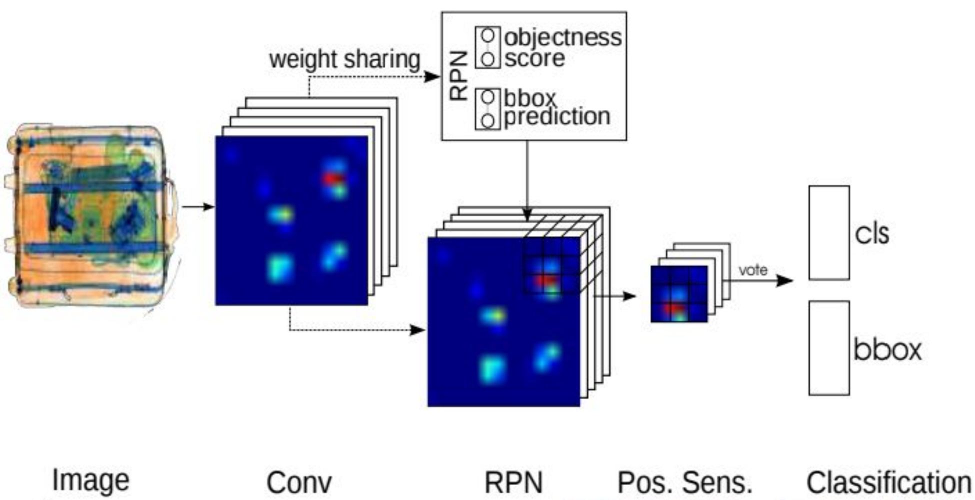
SW-CNN :



Faster RCNN:



R-FCN :



- Neural Network Provides Better Result than Traditional methods such as Bag of words etc.
- Examines the relative performance of contemporary object detection strategies for region based object detection techniques – SW-CNN , R-FCN and FRCNN. For Classification uses AlexNet, VGG{M,16} , ResNet-{50,101}.
- The Faster R-FCN and R-FCNN provide superior results than traditional sliding window driven CNN(SW-CNN) approach.

Model	Network	mAP	camera	laptop	gun	gun component	knife	ceramic knife
SWCNN	AlexNet	60.8	68.2	60.9	74.8	71.4	21.2	68.3
	VGGM	63.4	70.7	63.7	76.3	73.1	24.6	71.9
	VGG16	64.9	70.1	72.4	75.2	75.7	22.3	73.4
	ResNet-50	67.1	69.2	80.1	74.7	76.1	31.4	71.3
	ResNet-101	77.6	88.1	90.2	83.1	84.8	39.2	80.3
FRCNN	AlexNet	78.8	89.3	75.6	91.4	87.4	46.7	82.3
	VGGM	82.3	90.0	83.4	91.8	87.5	54.2	86.9
	VGG16	**88.3**	88.1	91.8	92.7	93.8	72.1	91.2
	ResNet-50	85.1	84.4	87.9	91.6	90.1	67.7	88.9
	ResNet-101	87.4	85.7	90.4	93.1	91.1	73.2	90.7
RFCN	ResNet-50	84.6	89.4	92.8	93.2	91.8	50.6	89.6
	ResNet-101	85.6	88.7	90.6	94.2	92.5	55.6	92.0

TRANSFER LEARNING USING CONVOLUTIONAL NEURAL NETWORKS FOR OBJECT CLASSIFICATION WITHIN X-RAY BAGGAGE SECURITY IMAGERY (2016)

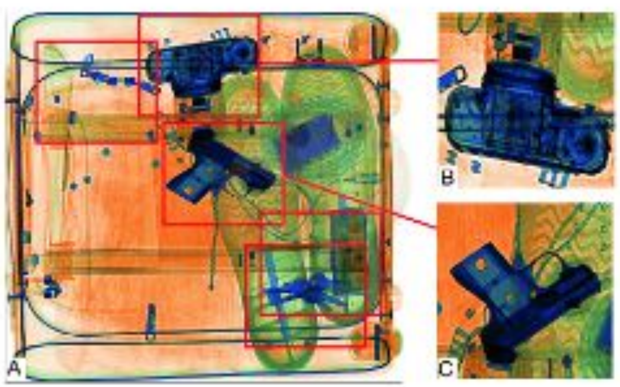


Fig. 2: Exemplar X-ray baggage image (A) with extracted data set regions for camera (B) and firearm (C) objects.

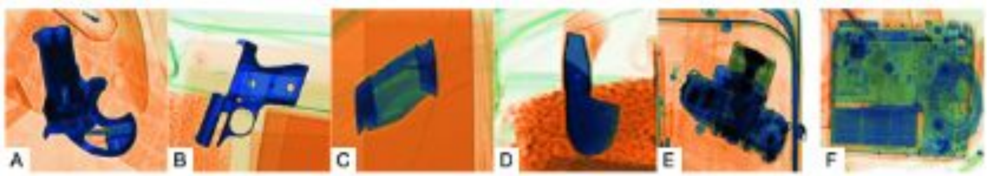


Fig. 3: Bag of visual words approach for multi-class problem. Type of baggage objects and the number of samples in our dataset is as follows: (A) Guns, (B) Gun Components, (C) Knives, (D) Ceramic Knives, (E) Cameras, (F) Laptops

- **AlexNet vs SVM and RF**

	TP%	TN%	FP%	FN%	PRE	REC	ACC
$AlexNet_{1-8}$	**99.26**	95.92	4.08	**0.74**	0.74	**0.99**	0.96
$AlexNet_{2-8}$	98.53	97.60	2.40	1.47	0.83	**0.99**	**0.98**
$AlexNet_{3-8}$	96.32	**97.81**	**2.19**	3.68	0.84	0.96	0.98
$AlexNet_{4-8}$	95.59	97.04	2.96	4.41	0.79	0.96	0.97
$AlexNet_{5-8}$	98.16	95.32	4.68	1.84	0.71	0.98	0.96
$AlexNet_{6-8}$	96.32	94.85	5.15	3.68	0.69	0.96	0.95
$AlexNet_{7-8}$	94.49	96.35	3.65	5.51	0.75	0.95	0.96
$AlexNet_{8}$	95.22	95.79	4.21	4.78	0.73	0.95	0.96
$SURF + RF$	80.74	67.28	32.72	19.26	0.95	0.81	0.79
$SURF + SVM$ 6	85.81	88.24	11.76	14.19	**0.98**	0.86	0.86

Table 2: Performance for the two class problem using test set.

- **GoogleNet**

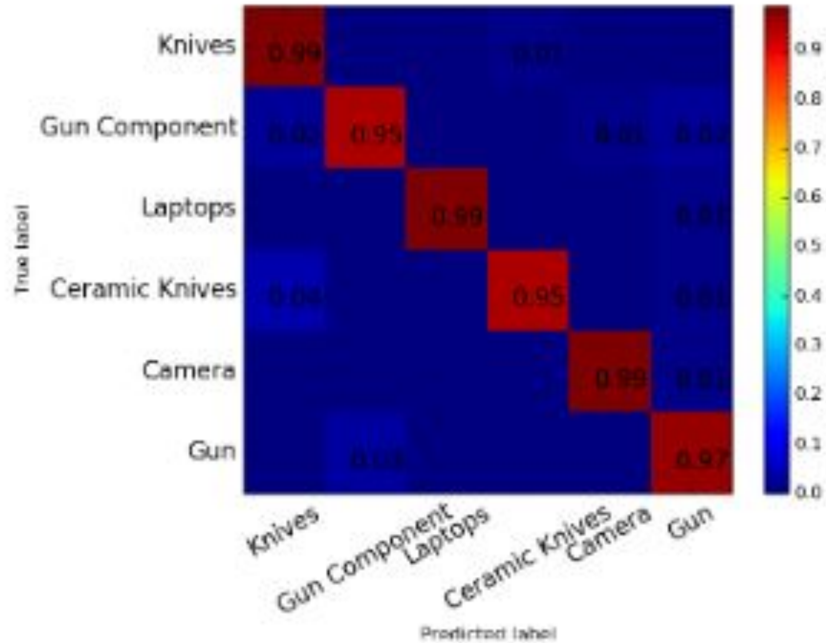


Fig. 4: Normalized confusion matrix of the fine-tuned GoogLeNet model tested on unseen test dataset.

	Camera	Laptop	Gun	Gun Component	Knives	Ceramic Knives	mAP
AlexNet	97.23	99.70	97.30	89.64	93.19	94.50	95.26
GoogLeNet	97.14	92.56	99.50	97.70	95.50	98.40	**98.40**

Table 3: Results for the multi-class problem (average precision %).

Automated Deep Learning for Threat Detection in Luggage from X-ray Images (2019)

Figure 2 A sample image containing a steel barrel bores (top right cylinder in the top row) from the parcel dataset. The left image (both rows) is the raw dual view x-ray scan, in the middle, the grey scale smoothed one, and on the right, b/w thresholded one. The parcel dataset usually contains a higher amount of steel objects and the barrels are better concealed.

- **Techniques: CNN and Stacked AutoEncoders with Shallow NN and Random Forest**
- **F1-score metric**

CNN	87%
Shallow NN	63%,
Random Forest	58%
Stacked Autoencoders	55%

The Good, the Bad and the Ugly: Evaluating Convolutional Neural Networks for Prohibited Item Detection Using Real and Synthetically Composited X-ray Imagery

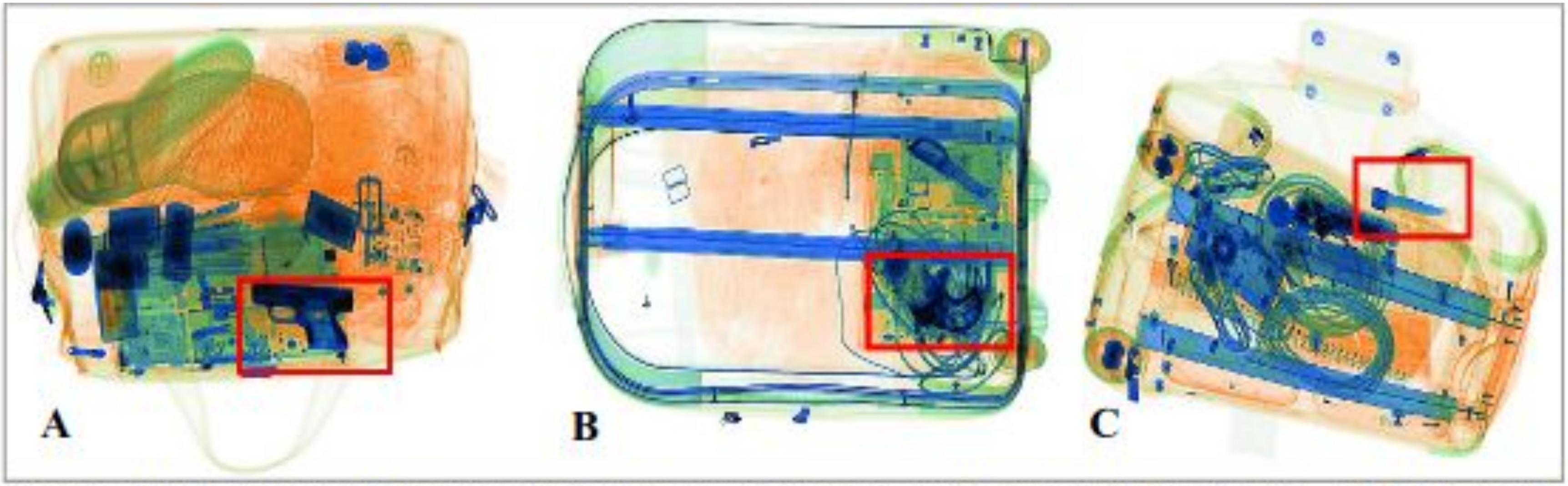


Figure 1: Exemplar X-ray security baggage images with prohibited objects - red box: (A) Firearm (B) Firearm Parts and (C) Knife.

UNIVERSITY OF ALBERTA

Threat Image Projection (TIP).

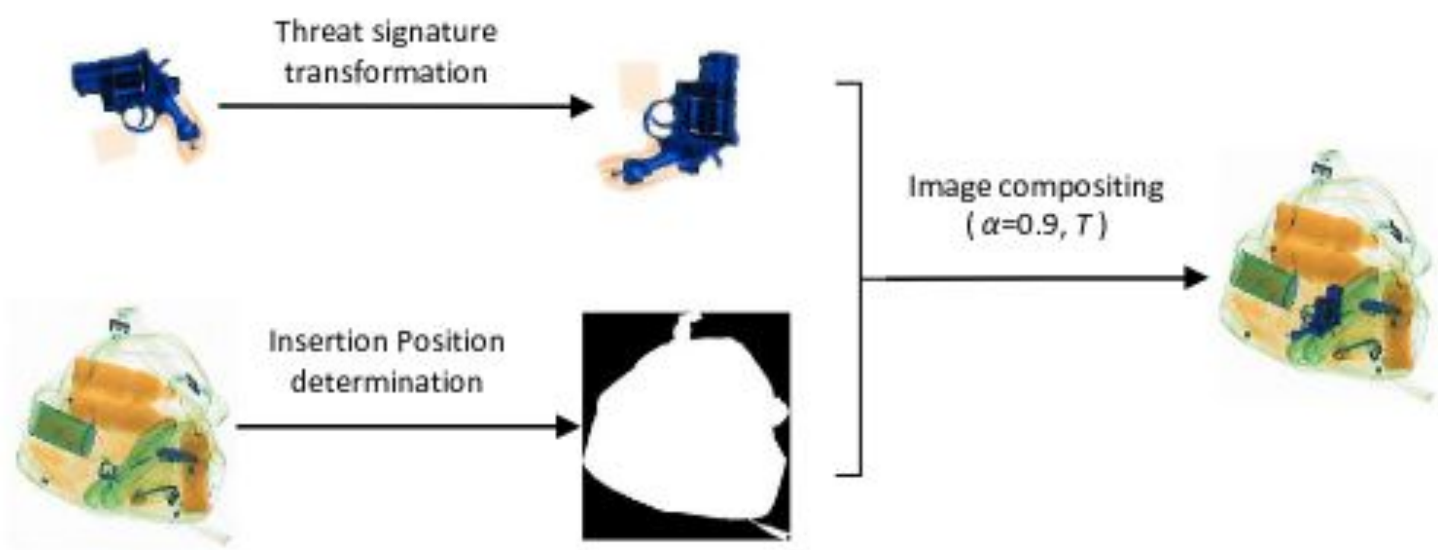


Figure 2: Threat image projection (TIP) pipeline for synthetically composited image generation.

Example of Generated Image

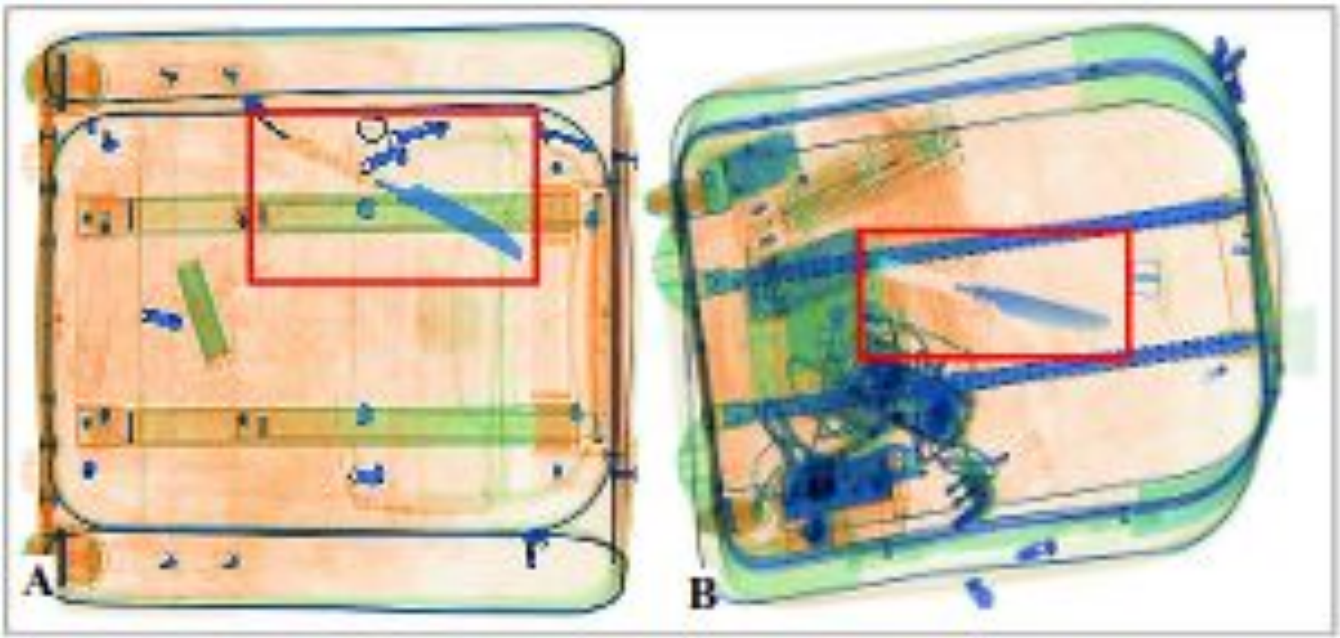


Figure 4: Visual comparison of real (A) and SC (B) X-ray security imagery of prohibited items.

Results on Real vs Synthetic DataSet

Train ⇒ Evaluation	Model	Network	Average precision			mAP
			Firearm	Firearm Parts	Knives	
$Dbf3_{Real} \Rightarrow Dbf3_{Real}$	Faster R-CNN [21]	$ResNet_{50}$	0.87	0.84	0.76	0.82
		$ResNet_{101}$	**0.91**	**0.88**	**0.85**	**0.88**
	RetinaNet [12]	$ResNet_{50}$	0.88	0.86	0.73	0.82
		$ResNet_{101}$	0.89	0.86	0.73	0.83
$Dbf3_{SC} \Rightarrow Dbf3_{Real}$	Faster R-CNN [21]	$ResNet_{50}$	0.82	0.77	0.55	0.71
		$ResNet_{101}$	**0.86**	**0.80**	**0.66**	**0.78**
	RetinaNet [12]	$ResNet_{50}$	0.84	0.77	0.53	0.71
		$ResNet_{101}$	0.84	0.76	0.54	0.72
$Dbf3_{Real+SC} \Rightarrow Dbf3_{Real}$	Faster R-CNN [21]	$ResNet_{50}$	0.85	0.79	0.65	0.76
		$ResNet_{101}$	**0.87**	**0.81**	**0.74**	**0.81**
	RetinaNet [12]	$ResNet_{50}$	0.85	0.81	0.64	0.76
		$ResNet_{101}$	0.86	0.80	0.63	0.76

Table 1: Detection results of varying CNN architecture trained on: Upper $\rightarrow$ $Dbf3_{Real}$, Middle $\rightarrow$ $Dbf3_{SC}$ and Lower $\rightarrow$ $Dbf3_{Real+SC}$. All models are evaluated on set of real X-ray security imagery.

Train ⇒ Evaluation	Model	Network	Average precision			mAP
			Firearm	Firearm Parts	Knives	
Dbf3$_{Real}$ ⇒ *Dbf3*$_{SC}$	Faster R-CNN [21]	ResNet_{50}	0.88	0.87	0.84	0.87
		ResNet_{101}	**0.92**	**0.92**	**0.89**	**0.91**
	RetinaNet [12]	ResNet_{50}	0.89	0.87	0.83	0.86
		ResNet_{101}	0.90	0.88	0.85	0.88
Dbf3$_{SC}$ ⇒ *Dbf3*$_{SC}$	Faster R-CNN [21]	ResNet_{50}	0.90	0.88	0.83	0.87
		ResNet_{101}	**0.93**	**0.92**	**0.86**	**0.91**
	RetinaNet [12]	ResNet_{50}	0.91	0.89	0.84	0.88
		ResNet_{101}	0.91	0.89	0.83	0.86
Dbf3$_{Real+SC}$ ⇒ *Dbf3*$_{SC}$	Faster R-CNN [21]	ResNet_{50}	0.89	0. 86	0.83	0.86
		ResNet_{101}	**0.91**	**0.89**	**0.87**	**0.89**
	RetinaNet [12]	ResNet_{50}	0.90	0.87	0.83	0.87
		ResNet_{101}	0.90	0.88	0.84	0.87

Table 2: Detection results of different CNN architecture trained on: Upper → *Dbf3*$_{Real}$, Middle → *Dbf3*$_{SC}$ and Lower → *Dbf3*$_{Real+SC}$. All models are evaluated on set of SC dataset.

Meta-Transfer Learning Driven Tensor-Shot Detector for the Autonomous Localization and Recognition of Concealed Baggage Threats

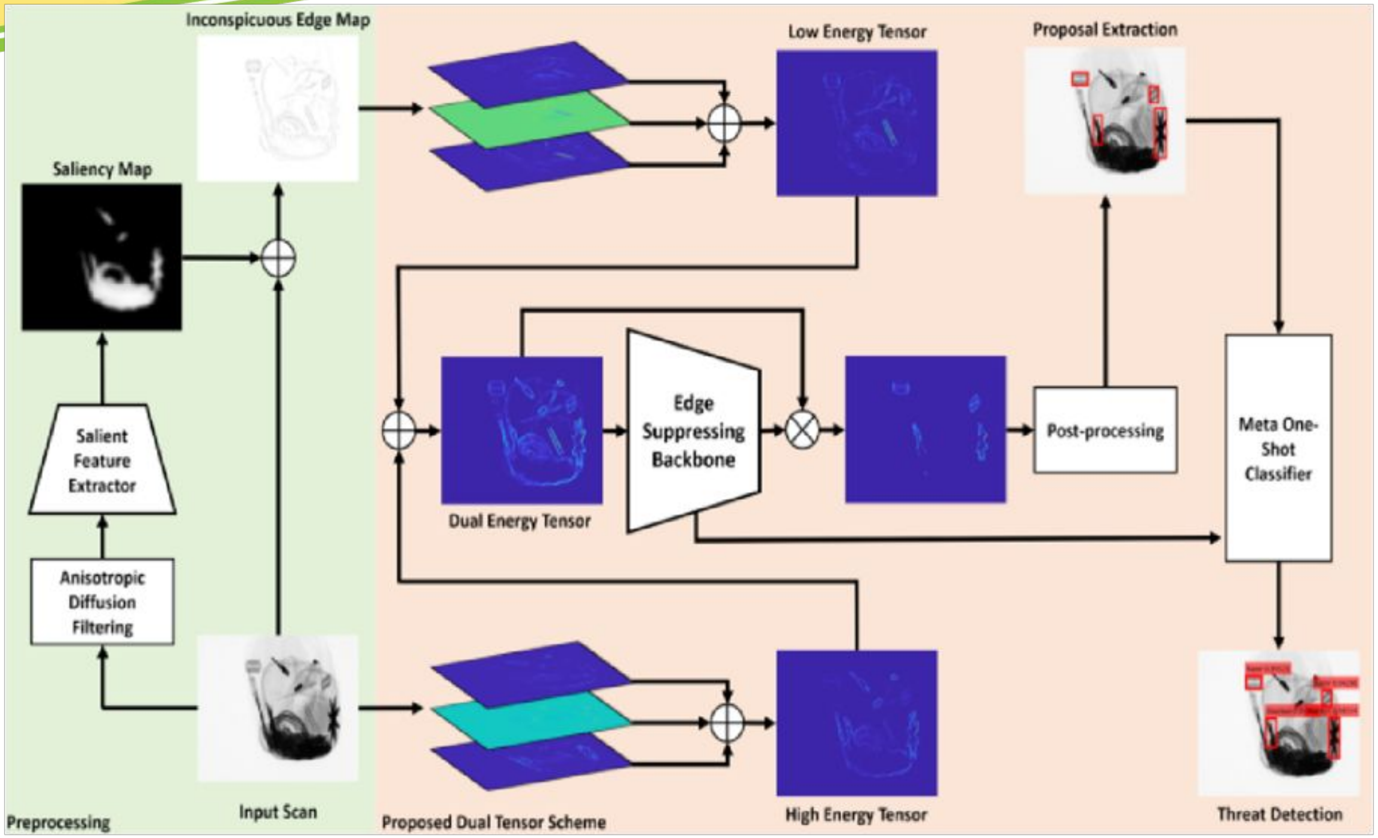


Figure 2. Block diagram of the proposed tensor-shot detector

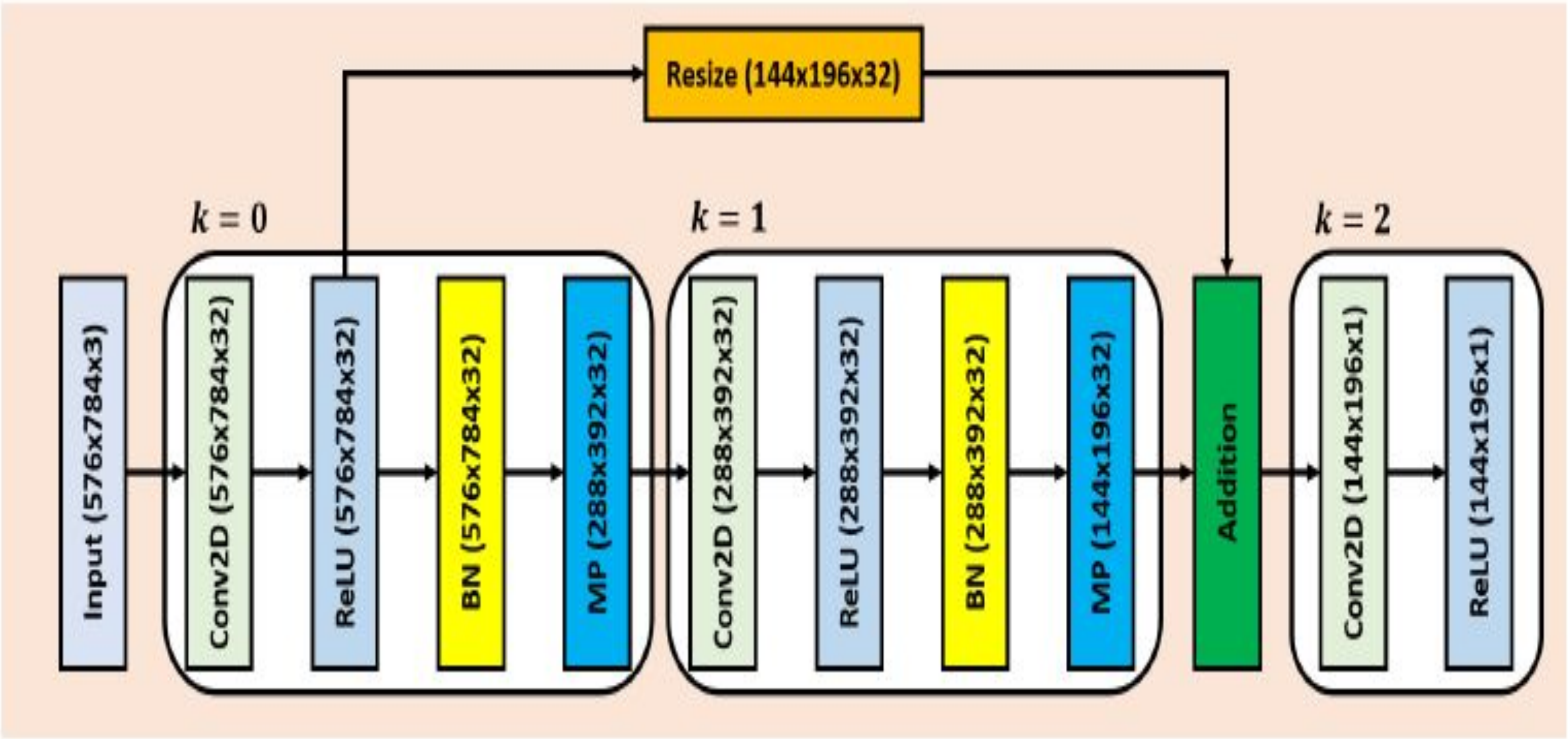


Figure 3. Salient network architecture. The abbreviations are Conv2D: 2D Convolution, BN: Batch Normalization, and MP: Max Pooling.

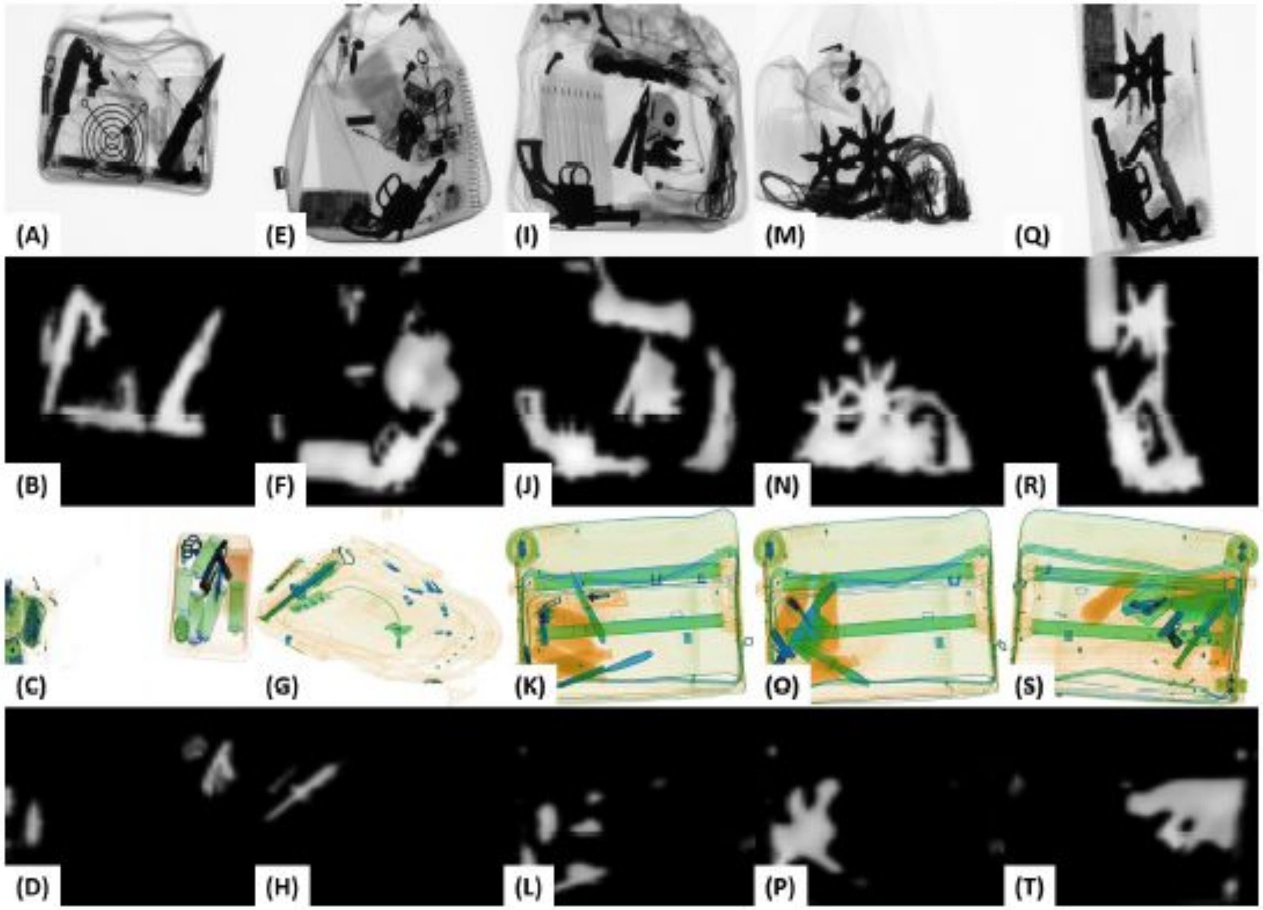


Figure 6. Saliency maps produced by the proposed salient feature extractor on both GDXray [15] and SIXray [14] datasets are shown in (**A–T**). Also, the first and third row show the original scans.

- Novel meta-transfer learning based single-shot detector capable of recognizing baggage threats under extreme occlusion

- A highly generalizable detection framework that leverages the proposed dual-tensor scheme to localize and recognize the threatening items without retraining the backbone on large set of examples

- Generalized framework that leverages meta-transfer learning to autonomously recognize concealed baggage threat from joint datasets(GDXray & SIXray).

Questions?

References

[1] Samet Akc ̧ay, Mikolaj E Kundegorski, Michael Devereux, and Toby PBreckon. Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery. In 2016 IEEE International Conference on Image Processing (ICIP), pages 1057–1061. IEEE, 2016.
[2] Muhammet Bas ̧tan, Mohammad Reza Yousefi, and Thomas M Breuel. Visual words on baggage x-ray images. In International Conference on Computer Analysis of Images and Patterns, pages 360–368. Springer, 2011.
[3] Neelanjan Bhowmik, Qian Wang, Yona Falinie A Gaus, Marcin Szarek, and Toby P Breckon. The good, the bad and the ugly: Evaluating convolutional neural networks for prohibited item detection using real and synthetically composited x-ray imagery. arXiv preprintarXiv:1909.11508, 2019.
[4] Victoria Cutler and Susan Paddock. Use of threat image projection (tip) to enhance security performance. In 43rd Annual 2009 International Carnahan Conference on Security Technology, pages 46–51. IEEE, 2009.

[5] Taimur Hassan, Muhammad Shafay, Samet Akçay, Salman Khan,Mohammed Bennamoun, Ernesto Damiani, and Naoufel Werghi. Meta-transfer learning driven tensor-shot detector for the autonomous localization and recognition of concealed baggage threats. Sensors, 20(22):6450, 2020.
[6] Domingo Mery, Vladimir Riffo, Uwe Zscherpel, German Mondragón, Iván Lillo, Irene Zuccar, Hans Lobel, and Miguel Carrasco. Gdxray: The database of x-ray images for nondestructive testing. Journal of Nondestructive Evaluation, 34(4):1–12, 2015.
[7] 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 Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2119–2128, 2019.
[8] Stefan Michel, Saskia M Koller, Jaap C de Ruiter, Robert Moerland, Maarten Hogervorst, and Adrian Schwaninger. Computer-based training increases efficiency in x-ray image interpretation by aviation security screeners. In 2007 41st Annual IEEE international Carnahan conference on security technology, pages 201–206. IEEE, 2007.
[9] Alessio Petrozziello and Ivan Jordanov. Automated deep learning for threat detection in luggage from x-ray images. In International Symposium on Experimental Algorithms, pages 505–512. Springer, 2019.
[10] 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.
[11] 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