Acta Scientific COMPUTER SCIENCES

Volume 4 Issue 8 August 2022

Research Article

Multistage 3D Shape Detection and Classification Network for Threat Items in CT Volumes of Scanned Luggage

Mohamed N Ahmed* and Fayin Li

IBM Fellow, IBM, India

***Corresponding Author:** Mohamed N Ahmed, IBM Fellow, IBM, India.

Received: April 25, 2022

Published: July 18, 2022

Abstract

In this work, we present a novel 2-stage system to detect and classify potentially hazardous objects in CT scans of carry-on luggage. The classification and detection approach consists of two 3D neural networks: Region proposal network (RPN) followed by a 3D shape classification network (SCN). RPN segment an input volume into 2 classes: Threat and background. To reduce the number of false positive regions identified by RPN, connected components labeling and various morphological operation are then applied to filter proposed regions for second stage 3D shape classification using SCN. Experimental results show the effectiveness of the proposed system in detecting various threat objects with high detection rates, while producing low false positives.

Keywords: Deep Learning; Convolution Neural Networks; 3D Convolution; Segmentation; Object Detection and Classification

Introduction

X-ray type technologies have been used for airport security checks for several decades. Heightened regard for the detection of complex articles within baggage and parcels for air transit and other forms of transportation has led to an increased interest in the use of automatic recognition strategies. Items of interest can be however difficult to detect within this environment due to a range of orientation, clutter, and density confusion in a traditional two-dimensional (2D) X-ray projection. Object occlusion is a limitation of 2D X-ray scanners, which makes detection (automatically or by human operators) particularly challenging.

To address these issues, there has been an expansion of the use of Computed Tomography (CT) volumetric imagery, where a three-dimensional (3D) "voxel" representation of the baggage or personal item is reconstructed. Modern imaging technology uses multiple energy bands in these CT scanners, for enhanced materials discrimination, and can achieve real time scanning rates for bags in airport baggage/parcel handling operations by allowing screeners to quickly identify prohibited items without divestiture of electronics or liquids.

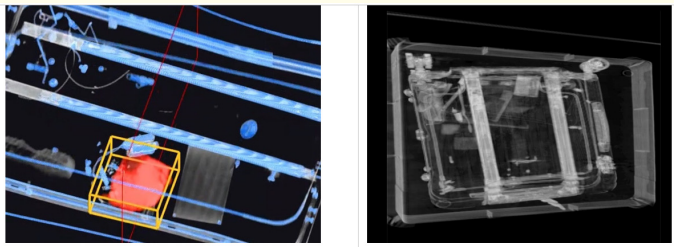

Figure 1: Examples of 3D CT scans of carry-on luggage. CT scans overcome issues with orientation, clutter, and density confusion often present in traditional 2D X-ray scanning technology, while providing enhanced materials discrimination, and achieving real time scanning rates.

There is little published work around automatic recognition and classification of items in scanned baggage. Most of this work revolves around using engineered features (color, edges, histograms, etc.) and classical classifiers (Support Vector Machines, Statistical classifiers, and neural networks). More recently, deep learning techniques have emerged as a powerful alternative for supervised learning with great model capacity and the ability to learn highly discriminative features for the task at hand. These features often outperform hand-crafted and pre-defined feature sets. Convolutional Neural Networks (CNNs) have been applied with promising results on a variety of imaging problems. The network topology exploits the stationary nature of natural images by learning features using locally connected networks. Convolutional layers are used to learn small feature detectors based on patches randomly sampled from a large image, resulting in powerful object representations without the need to hand design features. This has been empirically demonstrated on the challenging ImageNet classification task across thousands of classes.

Though 2D CNN can learn the spatial relationship between the pixels in a 2D plane, it cannot learn the inter-slice relationship between the frames in a 3D volume. To address this issue, 3D CNNs extends 2D CNNs by using convolution kernels in three dimensions, and hence the inter-slice information can be learned, providing better segmentation and classification results.

3D CNN-based architectures have been used extensively in medical image analysis [1-18], and in this work, we customize some of these architectures to identify potentially hazardous objects in scanned luggage.

Main contributions

To our knowledge, this is a first deep learning-based approach to detect and classify threat object in CT scans. Our approach and network architecture aims to strike the balance between learning powerful and discriminative 3D shape and material representation while providing the most optimized architecture complexity to meet required execution time for operational deployment.

Material and Methods

As illustrated in figure 2, the classification and detection approach consists of two 3D neural networks: Region proposal network (RPN) followed by a 3D shape classification network (SCN). RPN segment an input volume into 2 classes: Threat and background. To reduce the number of false positive regions identified by RPN, connected components labeling and various morphological operation are then applied to filter proposed regions for second stage 3D shape classification using SCN.

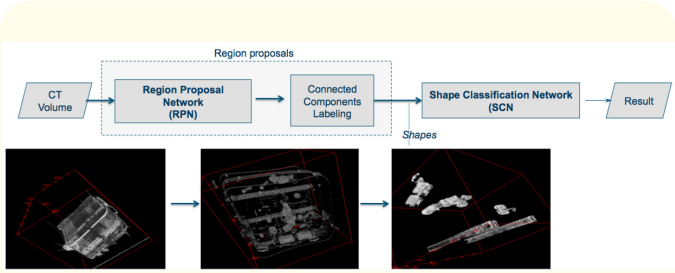

Figure 2: System Overview. The proposed 2-stage detection and classification approach consists of 2 3D neural networks: Region proposal network (RPN) followed by a 3D shape classification network (SCN).

Region proposal network (RPN)

RPN is trained to output a segmentation map of the central cube of a given sub-volume within the scanned bag. To explicitly capture spatial relationships in a three-dimensional context, we apply 3D convolutions on sub-volumes of different sizes. RPN is an 8-layers deep network with residual connections. Residual connections were shown to facilitate preservation of the flowing signal and as such have enabled training of very deep neural networks increasing the number of trainable parameters. Each block (L) consists of 2 convolution layers with small 3 x 3 x 3 convolution kernels. The use of small convolution kernels, and its small computation requirements, allows for building high performing yet efficient 3D CNNs. The input to the network is a sub volume of 25 x 25 x25 voxels and the output is a 9 x 9 x 9 classification map of the centroid within the sub volume.

Shape classification network (SCN)

To reduce the number of false positive regions identified by RPN, connected components labeling and various morphological operation are then applied to filter proposed regions for second stage 3D shape classification using SCN. Similar to RPN architecture, SCN is a 7-layers deep network with residual connections. The network

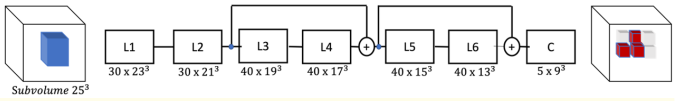

Figure 3: RPN architecture with residual connections. The operations within each layer block (L) are applied in the order: Batch-Normalization, non-linearity, and 3D convolution. C denotes classification layer.

consists of a series of convolution blocks and RLU layers, followed by fully connected layers for classification. Each convolution block (L) consists of 2 convolution layers with 3 x 3 x 3 convolution kernels. SCN is trained using threat shapes and background volumes.

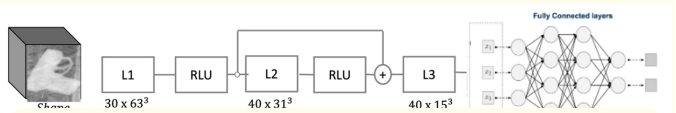

Figure 4: 3D Shape Classification network (SCN).

To validate the effectiveness of our method, the system was trained using 5000 CT scans containing various threat objects including guns and sharps of various types, sizes, and shapes, and 5000 of clear stream of commerce (SOC) CT scans. Each CT scan has 640 x 480 x 220 voxels, with a spatial resolution of 0.5–0.625 mm in the z-axis, and approximately 0.5 mm in the x- to y-axes. Figure 5 illustrates a sample of threat objects included in the training set.

Preprocessing and augmentation

Training datasets were augmented using scaling, 3D rotations of different angles, and various image artifacts. This process resulted in 300000 training and validation shapes.

Training details

For RPN training, we created the training and validation sets by sampling all threat voxels across every CT volume to form 50% of the samples. Non-threat voxels were sampled from a thresholded range to form 45% of the samples. The thresholds were computed by computing the mean of all threat voxels for a specific threat and subtracting 3.5 standard deviations from the mean for the low threshold and adding 3.5 standard deviations from the mean for the high threshold. The last 5% of the samples were randomly sampled from any non-threat voxel in the CT volume. Finally, 80% of the samples were used for training and 20% of the examples were used for validation. Training took approximately 76 seconds per epoch on a single NVIDIA Titan X (Maxwell generation) GPU with a batch size of 512 on approximately 2 million examples.

For SCN training, cropped 3D shapes and background sub-volumes were extracted from the training set. Training took approximately 400 seconds per epoch on a single NVIDIA Titan X (Maxwell generation) GPU.

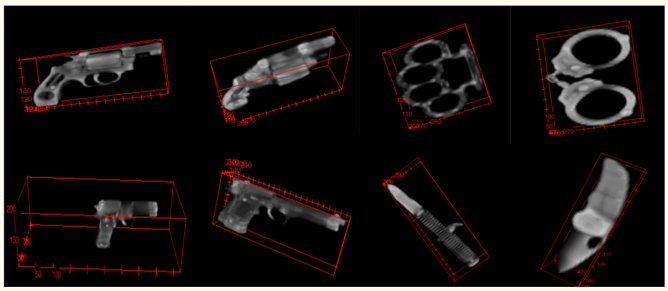

Figure 5: Examples of prohibited items used in data augmentation.

Evaluation metric

For quantitative performance evaluation, we use Intersection over union (IOU) metric using the model output predicted mask compared to the ground truth mask.

Where and are the output and ground truth mask, respectively.

Testing

For testing, a test set of 30,000 CT scans of various threat objects including guns and sharps of various types, sizes, and shapes,

as well as SOC CT scans obtained from different scanners were used. Tables 1-4 below, present testing results for 2 models: guns and sharps when applied to threat and clear SOC volumes. Across all CT scans, guns model has an 92% average detection rate with 8% false positives, while sharps model has 94% detection rate with 5% false positives

Datasets	Threat count	TP count	TP rate	FP count	FP rate	IOU rate (>= 50%)	IOU rate (>= 75%)	Inference Time(s)
G1	966	742	0.768	6	0.009	0.74	0.59	2.99
G2	964	741	0.769	6	0.009	0.752	0.695	2.804
G3	154	137	0.89	7	0.045	0.89	0.838	2.271
G4	41	25	0.61	0	0	0.585	0.341	2.586
G5	799	782	0.979	1	0.001	0.979	0.977	2.655
G6	799	779	0.975	1	0.001	0.975	0.975	2.671
G7	1657	1590	0.96	32	0.019	0.956	0.949	3.103
G8	1775	1728	0.974	1	0.001	0.973	0.973	2.9
Summary	7155	6524	0.912	54	0.008	0.905	0.873	2.876

Table 1: Results of testing on guns threat volumes.

Datasets	Count	FP count	FP rate	Inference Time
S1	608	1	0.002	2.261
S2	222	6	0.027	2.308
S3	627	1	0.002	2.251
S4	278	3	0.011	2.23
S5	2285	59	0.026	2.258
S6	2434	31	0.013	2.043
S7	1644	6	0.004	2.353
S8	852	1	0.001	2.324
S9	667	5	0.007	2.49
S10	554	10	0.018	2.397
S11	831	3	0.004	2.015
S12	1644	8	0.005	2.019
S13	602	3	0.005	2.168
Summary	13248	137	0.01	2.203

Table 2: Results of testing gun model on SOC volumes.

Datasets	Threat count	TP count	TP rate	FP count	FP rate	IOU rate (>= 50%)	IOU rate (>= 75%)	Inference Time(s)
K1	31	15	0.484	5	0.161	0.323	0.258	3.579
K2	202	171	0.847	12	0.059	0.782	0.688	2.867
K3	201	171	0.851	11	0.054	0.806	0.751	2.883
K4	70	57	0.814	3	0.044	0.8	0.743	2.226
K5	64	39	0.609	23	0.359	0.141	0.047	3.654
K6	956	923	0.965	4	0.004	0.963	0.837	3.544
K7	637	602	0.945	0	0	0.939	0.819	3.517
K8	999	960	0.961	9	0.009	0.959	0.835	3.444
K9	1007	967	0.96	9	0.009	0.96	0.831	3.358
K10	1682	1632	0.97	86	0.051	0.935	0.851	3.24
K11	679	660	0.972	39	0.057	0.934	0.897	3.324
K12	1279	1136	0.888	155	0.149	0.819	0.695	3
K13	460	439	0.954	19	0.041	0.904	0.904	3.357
K14	317	289	0.912	10	0.032	0.861	0.833	2.814
K15	335	305	0.91	4	0.012	0.86	0.836	2.902
Summary	8919	8366	0.938	389	0.045	0.905	0.811	3.266

Table 3: Results of testing sharps model on threat volumes.

Datasets	Count	FP count	FP rate	Inference Time
S1	608	8	0.013	2.654
S2	222	2	0.009	2.732
S3	627	13	0.021	2.656
S4	278	2	0.007	2.602
S5	2285	99	0.043	2.681
S6	2434	117	0.048	2.443
S7	1644	68	0.041	2.835
S8	852	29	0.034	2.736
S9	667	26	0.039	2.944
S10	554	13	0.023	2.789
S11	831	8	0.01	2.399
S12	1644	30	0.018	2.402
S13	602	34	0.056	2.579
Summary	13248	449	0.034	2.617

Table 4: Results of testing sharps model on SOC volumes.

Conclusions

In this work, we presented a novel 2-stage system to detect and classify potentially hazardous objects in CT scans of carry-on luggage. The classification and detection approach consists of two 3D neural networks: Region proposal network (RPN) followed by a 3D shape classification network (SCN). RPN segment an input volume into 2 classes: Threat and background. To reduce the number of false positive regions identified by RPN, connected components labeling and various morphological operation are then applied to filter proposed regions for second stage 3D shape classification using SCN. As shown in the above presented results, the system was capable of detecting various threat objects with high detection rates and low false positives, while meeting required execution times for operational application.

Bibliography

1. S P Singh., *et al*. "3D deep learning on medical images: a review". *Sensors* 20.18 (2020).

2. X Liu., *et al*. "A review of deep-learning-based medical image segmentation methods". *Sustainability* 13.3 (2021).

3. AM Rickmann., *et al*. "Recalibrating 3D convnets with project and excite". IEEE Trans. on medical imaging 39.7 (2020): 2461-2471.

4. S Niyas., *et al*. "Segmentation of focal cortical dysplasia lesions from magnetic resonance images using 3d convolutional neu-

ral networks". *Biomedical Signal Processing and Control* 70 (2021).

5. S Peng., *et al*. "Multi-scale 3d u-nets: An approach to automatic segmentation of brain tumor". *International Journal of Imaging Systems and Technology* (2020).

6. A Hatamizadeh., *et al*. "UNETR: Transformers for 3D Medical Image Segmentation". EEE/CVF Winter Conference on Applications of Computer Vision (WACV), 1748-1758, (2022).

7. Z Wu., *et al*. "Elnet: Automatic classification and segmentation for esophageal lesions using convolutional neural network". *Medical Image Analysis* 67 (2021).

8. Y Zhou., *et al*. "Multi-task learning for segmentation and classification of tumors in 3d automated breast ultrasound images". *Medical Image Analysis* 70 (2021).

9. X Zhu., *et al*. "Weakly supervised 3d semantic segmentation using cross-image consensus and inter-voxel affinity relations". Proceedings of the IEEE/CVF International Conference on Computer Vision (2021): 2834-2844.

10. Afonso., *et al*. "Towards Automated Melanoma Screening: Exploring Transfer Learning Schemes". *American Physical Society* (2016).

11. Asako., *et al*. "RotationNet: Joint Object Categorization and Pose Estimation Using Multiviews from Unsupervised Viewpoints". Mar. (2018).

12. Guo., *et al*. "Deep LOGISMOS: Deep Learning Graph-Based 3D Segmentation of Pancreatic Tumors on CT Scans". Jan. (2018).

13. Kamnitsas K., *et al*. "DeepMedic for Brain Tumor Segmentation". Proceedings of the International Workshop on (2016).

14. "Brain lesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries". Springer, Cham (2016): 138-149.

15. Kamnitsas K., *et al*. "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation". *Medical Image Analysis* 36 (2017): 61-78.

16. Litjens., *et al*. "A Survey on Deep Learning in Medical Image Analysis". *American Physical Society*, 4 June (2017).

17. R Holger., *et al*. "An Application of Cascaded 3D Fully Convolutional Networks for Medical Image Segmentation". *Computerized Medical Imaging and Graphics* 66 (2018): 90-99.

18. Suk HI., *et al*. "Deep Ensemble Learning of Sparse Regression Models for Brain Disease Diagnosis". *Current Neurology and Neuroscience Reports*., U.S. National Library of Medicine, Apr. (2017).

19. F Milletari., *et al*. "V-net: Fully convolutional neural networks for volumetric medical image segmentation". Fourth international conference on 3D vision (3DV), (2016): 565-571.