

Using deep learning on X-ray images to detect threats

Nicolas Jaccard¹, Thomas W. Rogers¹, Edward J. Morton², Lewis D. Griffin¹

¹University College London, ²Rapiscan Systems

Primary author contact email: n.jaccard@ucl.ac.uk

1 ABSTRACT

World trade volumes are exploding, with cargo containers totalling an estimated 500 million TEU (Twenty-foot equivalent units) shipped globally in 2012. At the same time, security requirements and transport regulations are increasingly stringent, putting significant pressure on the infrastructure at transport hubs and borders. In order to meet ambitious aims set by authorities, such as the inspection of every US-bound container, there is a pressing need to devise procedures to cope with high trade volumes while minimising the impact on the stream-of-commerce (SoC).

Non-invasive inspection systems such as X-ray instruments are routinely used to image the content of cargo containers. Imaging can be combined with risk analysis based on shipping manifest information (e.g. origin, destination) to selectively target containers that will undergo a costly and time-consuming physical inspection. However, the number of trained operators is insufficient to systematically assess all X-ray images acquired. Image processing algorithms are often embedded in X-ray inspection systems to enhance the image so that inspection times are reduced and accuracy improved. However, there is surprisingly little literature on the use of modern machine vision techniques for the *automated detection* of undeclared objects or threats in X-ray cargo images.

We have developed a deep learning framework for the classification of X-ray cargo images according to their content. This framework is based on convolutional neural networks (CNNs), a class of artificial neural networks that currently is the state-of-the-art in many areas of machine-learning and –vision. The performance of CNNs is better by a significant margin than previous methods and they have achieved better than human-level performance in some tasks, including photographic image classification. To our knowledge, CNNs have not yet been applied to X-ray cargo image classification.

CNNs typically require very large training datasets, the acquisition of which is costly and impractical for cargo images. However, the multiplicative nature of the X-ray transmission image formation process enables “sufficiently realistic” training images to be generated by projecting the objects to detect into SoC images.

In this contribution, we present an overview of our deep learning framework and present preliminary results, including a comparison to a more conventional approach we previously proposed for object detection in X-ray cargo container images. While our focus has been on cargo containers, we expect the framework to generalise to X-ray images of other vehicles and luggage. As such, this research is expected to contribute to the development of specialised software packages to assist operators through partial automation of the inspection workflow.

2 INTRODUCTION

It is estimated that the number of cargo containers shipped globally every year surpassed 500 million twenty-foot equivalent units (TEU) in 2012. As world trade volumes increase, so does the pressure on the security infrastructure at borders and transport hubs such as airports and ports. This pressure is further compounded by increasingly stringent regulations and policies set by governing bodies, such as an ambitious plan aiming at imaging all United States-bound cargo containers (Archick, 2010).

When crossing a security checkpoint, cargo containers are typically first subjected to a risk analysis based on the information provided in the shipping manifest such as countries of origin and destination or declared content (King, 2005; Weele and Ramirez-Marquez, 2010). The containers can also be imaged, either as a result of being flagged as suspicious by the risk analysis, or as part of a systematic imaging strategy. Cargo container imaging can be carried out using various non-invasive inspection systems, though X-ray transmission scanners are typically preferred where high energy radiation is required to penetrate through metallic walls (McDaniel et al., 2005). Both risk analysis and imaging are critical for the effective selection of suspicious containers that should undergo physical inspection, a costly and lengthy process that severely impedes the flow of commerce.

Machine vision techniques can be applied to cargo and luggage X-ray transmission images for image enhancement and automated analysis. Image enhancements, which represent the vast majority of such applications, aim at processing images in ways that facilitate their inspection by operators (Abidi et al., 2006, 2005; Ogorodnikov and Petrunin, 2002; Woodell et al., 2005). In contrast,

examples of automated analysis remain surprisingly rare in the literature (Mery, 2014) and mainly focus on X-ray images of baggage. For instance, support vector machines (SVM, a class of machine learning algorithms) and bag of words (BOW, a type of image descriptor) were used to detect threat objects in luggage images (Bastan et al., 2011; Flitton et al., 2015; Turcsany et al., 2013). Recently, automated analysis was applied to X-ray images of cargo containers, with the aim to classify their content into different categories (Tuszynski et al., 2013; Zhang et al., 2014). The performance reported by these studies remain relatively low, underlining the challenges associated with the automated classification of X-ray cargo images. Indeed, these images are noisy, cluttered, have a skewed perspective, and the objects they contain are transparent and can overlap (Bastan et al., 2011; McDaniel et al., 2005; Zhang et al., 2014). In addition, images can potentially contain any object that fits into a container, requiring automated schemes to suitably handle an extremely large variety of appearances and visual features.

We previously proposed an X-ray cargo image classification framework that was specifically developed to tackle these challenges (Jaccard et al., 2014a). Images were encoded using histograms of oriented Basic Image Features (oBIFs), a type of image feature conveying texture information that has been successfully applied to a large variety of machine vision problems (Crosier and Griffin, 2010; Jaccard et al., 2014b; Newell and Griffin, 2014). Classification of X-ray images based on histograms of oBIFs was learnt by an ensemble of random decision trees (Breiman, 2001), which typically perform as well or better than SVM while being much less demanding in terms of computational time. When applied to the detection of cars in cargo containers, this scheme delivered very high performance by detecting 100% of the cars, even those hidden behind other goods, with a false alarm rate lower than 1%. These encouraging results demonstrated the ability of such machine-learning schemes to perform well on X-ray cargo images.

In this contribution, we present an alternative approach for X-ray cargo images classification using convolutional neural networks (CNNs), a class of learning algorithm often referred to as deep learning (LeCun et al., 2015). In contrast with our previous scheme where a set of hand-crafted features (oBIFs) were used as image descriptors, convolutional neural networks learn the representation of the data (i.e. image features) that optimises its classification. CNNs are currently the state-of-the-art for many machine vision applications, outperforming other methods by a significant margin, including for large-scale photographic image classification where they were shown to exceed Human performance (He et al., 2015). Although CNNs have been successfully applied to biomedical X-ray images (Bar et al., 2015; Cernazanu-Glavan and Holban, 2013), we could

not find published examples of their application to either cargo or luggage X-ray transmission images.

We explore the use of CNNs for the classification of X-ray images of cargo containers. Despite novel regularisation techniques (Ioffe and Szegedy, 2015; Srivastava et al., 2014), CNNs still typically require very large datasets (e.g. 10^5 images) to prevent overfitting. Due to the monetary and time investment required for the acquisition of training X-ray cargo images, a method for the generation of synthetic images was devised. Preliminary classification results for small metallic threats are presented and compared to those obtained using our previous scheme based on hand-crafted features. This is a significantly more complex problem than car detection due to small threats not being restricted in size or orientation. Please note, we use the term 'small metallic threats' as we do not wish to make our research results easily discoverable by keyword searching. However, the objects in question are akin to small metallic objects (e.g. drill).

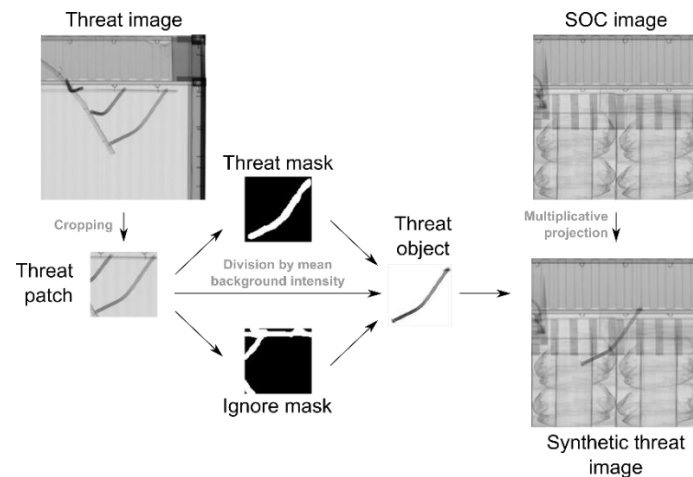
3 MATERIALS AND METHODS

3.1 Dataset and synthetic image generation

All stream-of-commerce (SoC) images used in this study were acquired using a Rapiscan®R60 rail scanner. Synthetic threat images were generated by projecting threat objects into SoC images, thus creating an arbitrarily large dataset with high variability in background appearances (Figure 1). This was made possible by the multiplicative nature of the X-ray transmission image formation process, whereby the attenuation of the beam by multiple objects is approximately the product of their individual attenuations (McDaniel et al., 2005).

A rectangular region of interest (ROI) encompassing the threat object to project was first defined by the user. A threat patch obtained by cropping the image was then manually annotated at a pixel level to produce two binary masks: a threat mask containing pixels corresponding to the threat, and an ignore mask that contains pixels of other non-background objects present in the threat patch. A rough background subtraction was then carried out by dividing the intensity of the pixels within the threat mask by the mean intensity of the threat patch pixels, ignoring pixels from the threat and ignore masks. Finally, the resulting threat object could be projected into a SOC image by multiplication.

Figure 1 Example of threat insertion into a stream-of-commerce (SoC) image. Note that this example does not depict an actual threat object used in this study for security reasons.



In total, approximately 700 threat objects spanning the range of possible orientations (e.g. multiple degrees of in-plane rotations) were manually segmented. 12,000 512×512 pixel background patches were randomly sampled from low to moderate density SOC images (similar to that shown in Figure 1). Before sampling, cargo container images were pre-processed as previously described (Jaccard et al., 2014a). For each background patch, the dataset comprised a negative class image (the background patch itself) and a positive class image (the background patch with the inserted threat), for a total of 24,000 images. Having both the negative and positive images in the dataset was empirically found to improve CNNs performance. The threat object to insert into a patch was randomly sampled, positioned, and flipped (vertically and horizontally). The intensity of the inserted object was also varied by a factor between -5% and 5% to introduce some additional noise. Images were downsampled by a factor of two to 256×256 pixel in order to minimise training and inference time.

About 20% of the dataset (≈ 4780 images across both classes) was used for testing and the rest for training. The training and testing sets were completely disjoint, with no overlaps in SOC backgrounds and threat objects inserted. The aim was to classify images in one of two classes (binary classification): negative class (no threat in image) and positive class (threat in image). Classification performance was evaluated by the way of receiver operating characteristic curves (ROC, plots of true positive rate versus false alarm rates). ROC curves were generated by thresholding the classifier output probability for a given patch to be a member of the threat (positive) class. The area-under-the-curve (AUC), as well as the recently introduced H-value (Hand, 2009) were used as metrics to quantitatively compare the performance of the approaches tested.

Table 3-1 Convolutional neural network architectures. Notation for convolutional layer is (**receptive field size – number of channels**) and for fully connected layers is (**number of channels**). Each convolutional layer and the first two fully connected layers are followed by a rectified linearity unit (ReLU, not shown for clarity). The two first fully connected layers are also followed by dropout (with probability set to 0.5).

CNN – 9 layers	CNN – 19 layers
Input (256×256 grayscale image)	
Convolutional layer (13-16)	Convolutional layer (3-8) Convolutional layer (3-8)
Max pooling	
Convolutional layer (11-32)	Convolutional layer (3-64) Convolutional layer (3-64)
Max pooling	
Convolutional layer (9-64)	Convolutional layer (3-128) Convolutional layer (3-128) Convolutional layer (3-128)
Max pooling	
Convolutional layer (7-128)	Convolutional layer (3-256) Convolutional layer (3-256) Convolutional layer (3-256)
Max pooling	
Convolutional layer (5-256)	Convolutional layer (3-512) Convolutional layer (3-512) Convolutional layer (3-512)
Max pooling	
Convolutional layer (3-512)	Convolutional layer (3-512) Convolutional layer (3-512) Convolutional layer (3-512)
Max pooling	
Fully connected layer (4096)	
Fully connected layer (4096)	
Fully connected layer (2)	
Soft max	

3.2 Convolutional neural network implementation

Convolutional neural networks (CNNs) were implemented using MatConvNet, an open source CNN implementation for MATLAB (Vedaldi and Lenc, 2015). A systematic evaluation of network architectures for X-ray image classification will be carried out in future work. However, two architectures stood out during preliminary experiments (see Table 3-1 for details). The first one is an architecture akin to early shallow models (Krizhevsky et al., 2012), with 9 layers. The second high performing architecture was a very deep network (Simonyan and Zisserman, 2014) with 19 layers. Both architectures have a similar number of parameters ($5 \cdot 10^7$ and $7 \cdot 10^7$ parameters for 9- and 19-

layer architectures, respectively) but differ significantly regarding the arrangement of their convolutional layers. The 9-layer CNN favours convolutional layers with large receptive fields (varying from 3 to 13 pixels) that are each followed by a pooling layer. In contrast, the 19-layer CNN only includes convolutional layers with small 3 pixels-wide receptive fields that are arranged in stacks of two or three in-between pooling layers. For both architectures, the learning rate was constant throughout training at $1 \cdot 10^{-3}$ while the weight decay for regularization was set to $2 \cdot 10^{-2}$ and $5 \cdot 10^{-4}$ for the 9- and 19-layers CNN, respectively. Training was carried out using an Nvidia Titan X GPU.

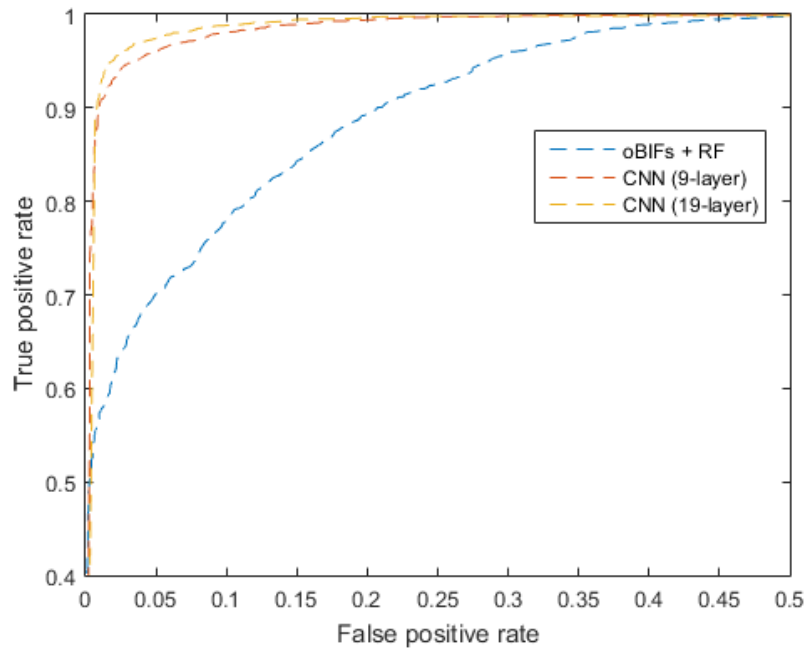
3.3 Classification using Basic Image Features and Random Forest classifiers

Classification performance using CNNs was compared to that obtained using our previous oBIFs and Random Forest scheme (Jaccard et al., 2014a). For each 256×256 grayscale image, oBIFs were computed at four different scales ($\sigma = 0.7, 1.4, 2.8, 5.6$) and for two threshold parameter values ($\epsilon = 0.011, 0.1$). The Random Forest classifier was made up of a thousand trees, with the *mtry* parameter (number of features randomly sampled at each split) set to 13 (i.e. the square root of the total number of features). Posterior probability was given by the fraction of trees assigning a given image to the threat (positive) class.

4 RESULTS

Threat X-Ray cargo image classification performance was evaluated for the CNN approaches (9- and 19-layer) as well as for our previous scheme based on oBIFs and random decision trees. Figure 2 shows the ROC curves obtained for these three approaches while summary metrics are shown in Table 4-1.

Figure 2 ROC curves for threat X-ray image classification using CNNs (in orange and yellow) and our previous scheme (oBIFs + Random Forest, in blue). These curves were obtained using the testing dataset.



Whereas our previous scheme excelled at detecting cars in X-ray cargo images, its performance for the detection of small metallic threats was underwhelming. For a 90% true positive rate, which is a reasonable target for such systems, the oBIFs and Random Forest resulted in a false positive rate greater than 21%. In contrast, using CNNs resulted in less than 1% false alarms for the same true positive rate. Both the AUC and H-value also reflected the substantial improvement in classification performance delivered by deep learning and showed that the 19-layer CNN consistently outperformed the 9-layers architecture.

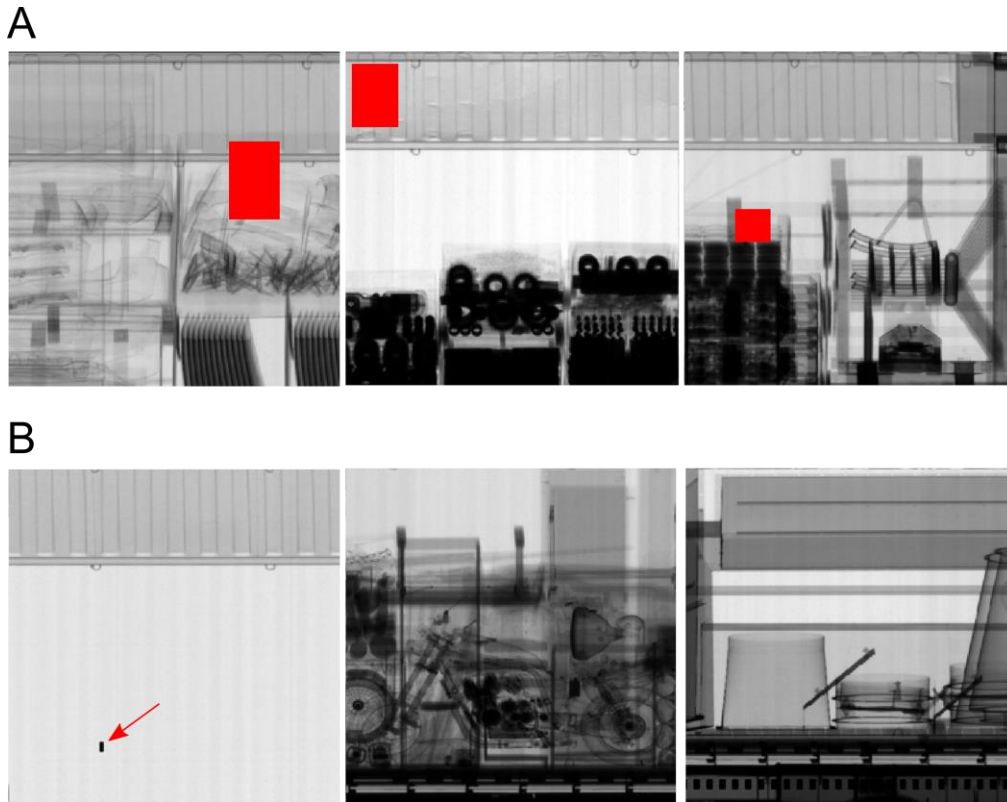
Table 4-1 Quantitative metrics for the comparison of our previous scheme (oBIFs + Random Forest) and CNN. AUC is the area under the ROC curve. PFR and TPR stand for false positive and true positive rates, respectively. These results were obtained using the testing dataset.

Method	AUC	H-value	FPR for 90% TPR [%]
oBIFs + Random Forest	0.9414	0.6252	21.02
CNN (9-layer)	0.9914	0.8868	0.92
CNN (19-layer)	0.9942	0.9101	0.80

The CNN approach successfully classified challenging images where threats were hidden amongst other goods or in the roof of the containers (Figure 3.A). Figure 3.B shows examples of false positive images (i.e. background images that were misclassified as containing threats). Many of them contain several objects with appearances similar to the inserted threats. It is also interesting to note that some images of empty containers were classified as containing a threat due to the presence of an

imaging artefact that resembled a small threat aligned with the X-ray beam's direction (i.e. front view of the threat).

Figure 3 Examples of true positive (A) and false positive (B) images after classification using the 19-layer CNN. The red boxes in (A) indicate the location of the inserted threat, the red arrow in (B) shows the location of an imaging artefact that was detected as threat by the CNN.



5 CONCLUSION

In this contribution, we presented a deep learning approach for the detection of small threats in X-ray cargo images. A suitably-large dataset of synthetic images was generated by exploiting the multiplicative nature of the X-ray transmission imaging formation process. The two convolutional neural networks (CNN) architectures considered (9- and 19-layers) significantly outperformed our previous X-ray cargo image classification scheme based on hand-craft image features and random decision trees. These results suggest that similarly to other machine vision applications, CNNs might become the state-of-the-art in automated X-ray cargo image analysis.

Future work will focus on a systematic evaluation of different CNN architectures and hyper-parameters. There will also be a need to thoroughly characterise the synthetic image generation process and compare its accuracy with real images of threats placed amongst other goods.

6 REFERENCES

- Abidi, B.R., Page, D.L., Abidi, M.A., 2005. A Combinational Approach to the Fusion, De-noising and Enhancement of Dual-Energy X-Ray Luggage Images, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops. IEEE, pp. 2–2. doi:10.1109/CVPR.2005.386
- Abidi, B.R., Zheng, Y., Gribok, A.V., Abidi, M.A., 2006. Improving Weapon Detection in Single Energy X-Ray Images Through Pseudocoloring. *IEEE Trans. Syst. Man Cybern. Part C (Applications Rev.* 36, 784–796. doi:10.1109/TSMCC.2005.855523
- Archick, K., 2010. U. S. -EU Cooperation Against Terrorism. DIANE Publishing.
- Bar, Y., Diamant, I., Wolf, L., Greenspan, H., 2015. Deep learning with non-medical training used for chest pathology identification, in: Hadjiiski, L.M., Tourassi, G.D. (Eds.), *SPIE Medical Imaging. International Society for Optics and Photonics*, p. 94140V. doi:10.1117/12.2083124
- Bastan, M., Yousefi, M.R., Breuel, T.M., 2011. Visual Words on Baggage X-Ray Images. pp. 360–368. doi:10.1007/978-3-642-23672-3_44
- Breiman, L., 2001. Random Forests. *Mach. Learn., Springer Series in Statistics* 45, 5–32. doi:10.1177/1062860609354639
- Cernazanu-Glavan, C., Holban, S., 2013. Segmentation of Bone Structure in X-ray Images using Convolutional Neural Network. *Adv. Electr. Comput. Eng.* 13, 87–94. doi:10.4316/AECE.2013.01015
- Crosier, M., Griffin, L., 2010. Using Basic Image Features for Texture Classification. *Int. J. Comput. Vis.* 88, 447–460.
- Flitton, G., Mouton, A., Breckon, T.P., 2015. Object classification in 3D baggage security computed tomography imagery using visual codebooks. *Pattern Recognit. In Press*, 1–11. doi:10.1016/j.patcog.2015.02.006
- Hand, D.J., 2009. Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Mach. Learn.* 77, 103–123. doi:10.1007/s10994-009-5119-5
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification.
- Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.
- Jaccard, N., Rogers, T.W., Griffin, L.D., 2014a. Automated detection of cars in transmission X-ray images of freight containers, in: 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, pp. 387–392. doi:10.1109/AVSS.2014.6918699
- Jaccard, N., Szita, N., Griffin, L.D., 2014b. Trainable segmentation of phase contrast microscopy

- images based on local Basic Image Features histograms, in: *Medical Image Understanding and Analysis*. British Machine Vision Association, pp. 47–52.
- King, J., 2005. The security of merchant shipping. *Mar. Policy* 29, 235–245. doi:10.1016/j.marpol.2004.04.006
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, in: *Advances in Neural Information Processing Systems*. pp. 1097–1105.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539
- McDaniel, F. Del, Doyle, B.L., Vizkelethy, G., Johnson, B.M., Sisterson, J.M., Chen, G., 2005. Understanding X-ray cargo imaging. *Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms* 241, 810–815.
- Mery, D., 2014. Computer vision technology for X-ray testing. *Insight - Non-Destructive Test. Cond. Monit.* 56, 147–155. doi:10.1784/insi.2014.56.3.147
- Newell, A.J., Griffin, L.D., 2014. Writer identification using oriented Basic Image Features and the Delta encoding. *Pattern Recognit.* 47, 2255–2265. doi:10.1016/j.patcog.2013.11.029
- Ogorodnikov, S., Petrunin, V., 2002. Processing of interlaced images in 4–10 MeV dual energy customs system for material recognition. *Phys. Rev. Spec. Top. - Accel. Beams* 5, 104701. doi:10.1103/PhysRevSTAB.5.104701
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., 2014. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15, 1929–1958.
- Tavasszy, L., Minderhoud, M., Perrin, J.-F., Notteboom, T., 2011. A strategic network choice model for global container flows: specification, estimation and application. *J. Transp. Geogr.* 19, 1163–1172. doi:10.1016/j.jtrangeo.2011.05.005
- Turcsany, D., Mouton, A., Breckon, T.P., 2013. Improving feature-based object recognition for X-ray baggage security screening using primed visualwords, in: *2013 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, pp. 1140–1145. doi:10.1109/ICIT.2013.6505833
- Tuszynski, J., Briggs, J.T., Kaufhold, J., 2013. A method for automatic manifest verification of container cargo using radiography images. *J. Transp. Secur.* 6, 339–356. doi:10.1007/s12198-013-0121-3
- Vedaldi, A., Lenc, K., 2015. MatConvNet: Convolutional Neural Networks for MATLAB 689–692. doi:10.1145/2733373.2807412
- Weele, S.F., Ramirez-Marquez, J.E., 2010. Optimization of container inspection strategy

via a genetic algorithm. *Ann. Oper. Res.* 187, 229–247. doi:10.1007/s10479-010-0701-6

Woodell, G., Rahman, Z., Jobson, D.J., Hines, G., 2005. Enhanced images for checked and carry-on baggage and cargo screening, in: Carapezza, E.M. (Ed.), *Third International Conference on Advances in Pattern Recognition, ICAPR 2005, Bath, UK, August 22-25, 2005, Proceedings, Part II*. doi:10.1007/11552499_1

Zhang, J., Zhang, L., Zhao, Z., Liu, Y., Gu, J., Li, Q., Zhang, D., 2014. Joint Shape and Texture Based X-Ray Cargo Image Classification, in: *2014 IEEE Conference on Computer Vision and Pattern Recognition Workshops*. IEEE, pp. 266–273. doi:10.1109/CVPRW.2014.48