Transferring X-ray based automated threat detection between scanners with different energies and resolution

Caldwell M, Ransley M, Rogers TW, and Griffin LD

Department of Computer Science, UCL, Gower Street, London, UK

ABSTRACT

A significant obstacle to developing high performance Deep Learning algorithms for Automated Threat Detection (ATD) in security X-ray imagery, is the difficulty of obtaining large training datasets. In our previous work, we circumvented this problem for ATD in cargo containers, using Threat Image Projection and data augmentation. In this work, we investigate whether data scarcity for other modalities, such as parcels and baggage, can be ameliorated by transforming data from one domain so that it approximates the appearance of another. We present an ontology of ATD datasets to assess where transfer learning may be applied. We define frameworks for transfer at the training and testing stages, and compare the results for both methods against ATD where a common data source is used for training and testing. Our results show very poor transfer, which we attribute to the difficulty of accurately matching the blur and contrast characteristics of different scanners.

Keywords: Deep Learning, X-ray imaging, Automated Threat Detection, Transfer Learning

1. INTRODUCTION

Air travel, maritime freight and parcel delivery have increasing volume in an increasingly security conscious world. Content screening via X-ray imaging is an essential component to the success and safety of all three modalities. However, whilst the throughput of imaging technologies has increased in step with the rising volume of transported items, human analysis of the resulting data remains a bottleneck, resulting in long and frustrating waits at airports and the majority of parcels and shipping containers going unchecked. Human operators are additionally susceptible to error, variability, scalability issues and potential corruption, so automation of this process is an attractive proposition for both industries and governments.

We and others have previously shown promising results applying deep learning methods adapted from the photographic domain to Automatic Threat Detection (ATD) within cargo containers.[1–3] The Convolutional Neural Networks (CNNs) which were used obtained accuracy by training on 10^5 labelled images generated through Threat Image Projection (TIP) onto a stream-of-commerce (SoC) dataset. It is easy to conceive of scenarios where, even with TIP, training data is limited in quantity and scope—for example when a new generation of scanner is introduced, or a new threat emerges—while producing additional training data would be costly and time consuming. This is compounded by secrecy around such data and materials within both state and commercial organisations.

Within the broad area of machine learning research many are attempting to address the dearth-of-data problem through *transfer learning*,[4] whereby networks trained on well formulated and densely populated domains can be put to use on ones where data is scarce. Previous work on transfer learning for ATD utilised an existing CNN trained on the extensive ImageNet dataset of optical photographs (10^6 images), which was then fine-tuned with a smaller dataset comprising X-rays of benign and threat-containing luggage (10^4 image patches).[5] This method obtained 98% accuracy at detecting handguns in luggage and found the fine-tuning to be most effective when restricted to the higher-level layers of the CNN. In other fields transfer learning has been found to speed up the training process, since low-level operations such as edge detection can be carried over by freezing initial layers whilst the remaining network is retrained. This approach has shown promising results in fields as diverse as character recognition,[6] tumour classification[7] and human pose estimation.[8] Other successful approaches

Further author information: (Send correspondence to L.D.G.)
L.D.G.: E-mail: l.griffin@cs.ucl.ac.uk

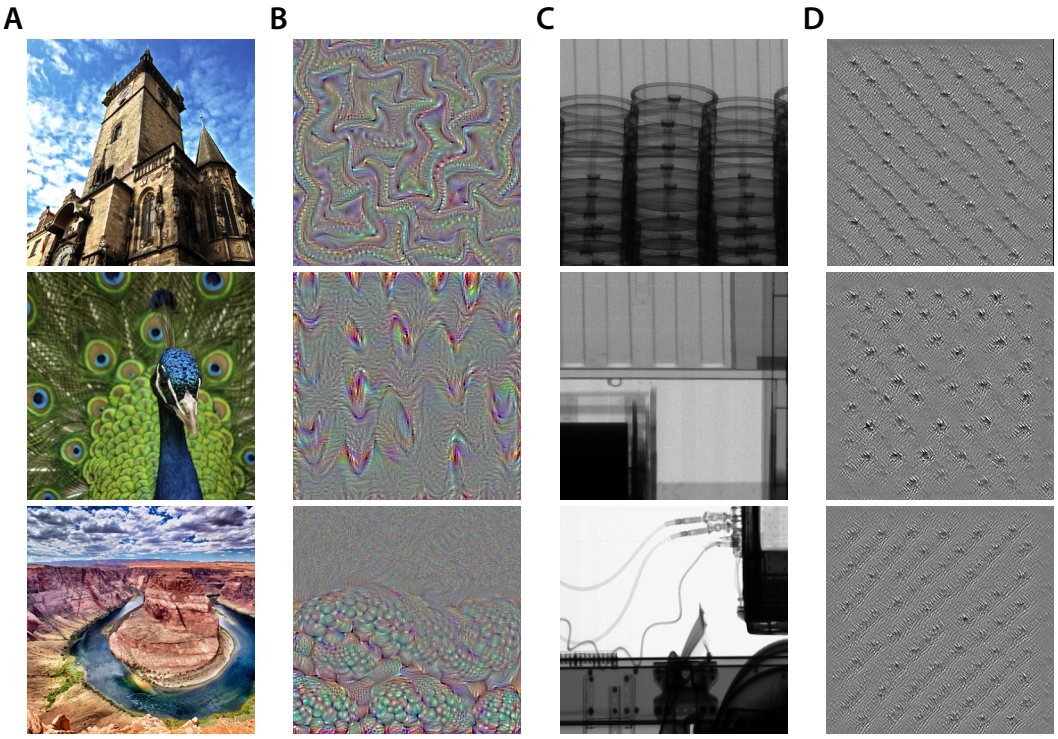

Figure 1. (**A**) Natural images such as those in ImageNet contain rich details that are important for classification into rich categories. (**B**) Deep networks for natural images must learn complex, hierarchical filters, such as those visualised here. (**C**) Security X-ray images are less varied and more stereotypical. Almost all detail that does occur can be present in both threat and non-threat images and hence is tangential to the binary classification task. (**D**) Filters learned for this task are simpler.

have included the use of stacked denoising auto-encoders to learn domain-invariant representations[9] and domain adaptation,[10] which is the approach we utilise here.

This paper reports our investigations into whether a CNN trained for ATD in one transport modality can be generalised to others. We start by presenting an ontology of X-ray images within security screening. This allows us to identify what types of change may be tackled by transfer learning. We then propose *training transfer* and *testing transfer* as two distinct frameworks for ATD, and evaluate the effectiveness of both approaches at ATD transfer between cargo and parcel data—modalities with differing resolution and colour profiles. Our results indicate that transforming data to comply with a pre-trained network is difficult, whilst using transformed data to train a network, though achievable, performs less well than when training and testing data are both drawn from a common source.

2. GENERALISABILITY IN DEEP LEARNING

Natural vision exhibits a very high level of generalisation. Humans can recognise an object they have never seen based on only a few, possibly very different, examples of the type—even rough drawings or verbal descriptions.[11] A typical 5 year old could probably recognise an undisguised weapon in an X-ray having seen only Han Solo's laser blaster and Yosemite Sam's cartoon pistol; this demonstrates a remarkable capacity for generalisation—both from specific objects to general forms and from one imaging modality to another—and constitutes an existence proof.

While it is currently unproven whether neural nets are capable of such deep conceptual transformations, convolutional networks such as VGG-19[12] do demonstrate better-than-human performance at classification over rich image categories in ImageNet.[13] These networks exhibit a remarkable degree of generalisability in terms of

recognising feature conjunctions in different permutations and contexts. The convolutional property—that the analytic filters are applied at every position—allows for translational invariance. Hierarchical structures of layers with spatial pooling allow for identification of feature relationships across different scales.

Such generalisability arises not purely from the CNN itself, but also from the structure of the problem space being learned. In the context of very large sets of natural images with a rich classification structure such as that provided in ImageNet, a CNN must learn very complex class boundaries and a rich set of features and relationships to support them. Visualisation methods such as activation maximisation[14] can be used to identify and illustrate the kinds of image features to which individual convolutional filters become attuned, revealing their complexity (Fig. 1B).

The space of X-ray security images is typically much less rich. This is in part because of the more formalised imaging environment: image content is more restricted and there is less variation in magnification (Fig. 1C). Perhaps more importantly, learning is constrained by the much lower granularity of classification. At the most basic level, we are interested only in distinguishing between threat and non-threat. Whereas contextual features can offer vital clues for conventional image classification, it seems likely that background details may often be peripheral to the ATD problem. That is not to say that non-threat features are irrelevant—for example, features of a threat confound such as an edge-on laptop may well be worth learning—but the space of relevant features may be smaller. Visualisation of the filters in a CNN trained on X-ray data does indeed suggest the network is selecting for less varied and complex features (Fig. 1D). Such features risk focussing too closely on small details that are very specific to the particular imaging context, or evolving case-specific "hacks", potentially limiting generalisation from one scanner to another. Limitations in the generalisability of ATD networks across modalities and scanners are important given the difficulty of obtaining high quality training data in the security industry. Building a CNN and training set from scratch for every variation on the domain would likely be commercially prohibitive, and seems unnecessary considering the high level of transferability humans exhibit in visual classification.

To summarise, CNNs do not necessarily generalise beyond the level that is required to solve the task they have been trained on, and aspects of the ATD problem may exacerbate this. The formulation of ATD as a binary classifier, and the generally limited and "hand made" datasets could present barriers to transferring good performance to novel scenarios. Additionally there is a high risk of overfitting to quirks of individual imaging systems—or worse, to synthesis artefacts which may not be present in SoC data. On the other hand we *do* want to exploit specifics where beneficial; e.g. subtle variations in dark regions of certain datasets that are not easily perceptible to the human eye. Hence this raises concerns that high generalisability could come at the price of losing sensitivity to subtle but consistent cues.

3. ONTOLOGY OF DATA FOR X-RAY THREAT DETECTION

Security X-ray imaging is used in many different contexts and at many different scales. In 2007 it was estimated that 5% of containers inbound to the USA were scanned.[15] An estimated 10.4 million items of airline cabin baggage are scanned every day worldwide, with a further 150,000 tonnes of air-freight.[16] The UK's Royal Mail handles over 1 billion parcels annually.[17] All figures are expected to rises as global trade increases and technical advances permit faster screening.[18] Each modality gives rise to images with very different properties and even within a specific transport modality there may be significant equipment differences. X-ray scanners operate at different energies and spatial resolutions. Some scanners provide basic greyscale imaging of X-ray transmission levels,[19] but many apply false-colouring algorithms to expose and highlight particular material properties.

The objects imaged in each modality are also very different from one another, and there may be differences in the relevant threats. For example not only are the types of benign payloads likely to differ between containers and parcels, along with their quantity and distribution, but the relevant threats are likely to be more compact if sent by parcel (e.g. handgun components rather than complete assault rifles).

Availability of data from different sources is highly variable and may be regulated by a range of actors at the commercial and governmental level. Several agencies have made efforts to explicitly generate useful datasets with representative examples of possible benign and threat objects. Large amounts of stream of commerce data

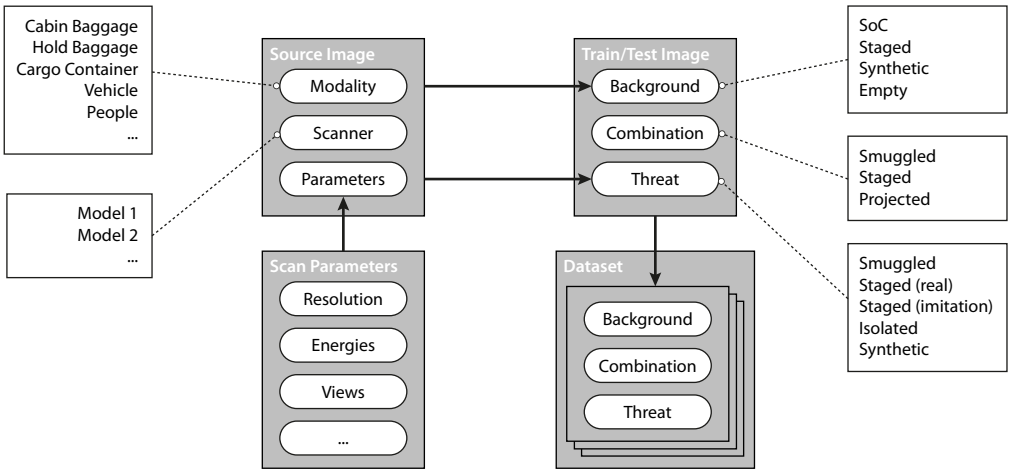

Figure 2. Ontology of data sources and types that may be employed for network training in the context of Automated Threat Detection (ATD). Data may be sourced from many different pieces of equipment, with different imaging parameters, deployed in different transport modalities. Imaged targets may be real stream of commerce (SoC) traffic, isolated objects, or staged constructs specifically devised for testing. A training dataset must contain both threat and benign images, potentially drawn from different sources and combined via techniques such as Threat Image Projection. A dataset contains many images, but the background and threat sources and method of combination remain fixed over the whole set. We can thus specify a dataset in terms of these three properties.

may be available in specific channels; this will usually have limited information about content, though in most cases can be assumed to be mostly benign.

In order to formalise what comprises a dataset suitable for ATD training, we conceptualise the data space as illustrated in Fig. 2.

A *source image* is characterised by the transport *modality* it is drawn from or represents, the *scanner* or other imaging equipment used to generate it, and an associated collection of scan *parameters*. Modalities include: aviation cabin baggage, aviation hold baggage, aviation freight, fast parcels, container cargo, pallet cargo, cars, trucks, and so on. Scanners come in a variety of configurations from a range of manufacturers; in addition, training images may be generated by simulation rather than scanning. Imaging parameters include:

- spatial resolution: pixel extent and point-spread-function width
- bit-depth and effective precision: scanners may be quoted as 16-bit but only 12 may be reliable
- X-ray energies: spectral peak and width
- viewing angles: single- dual- and quad-view are all available, with variations in the relative orientation of the views
- image-sensor transfer functions: intensity nonlinearities and false colouring

A *training/test image* may directly correspond to a single source image, or it may be constructed from multiple sources. Real smuggled threats occur very infrequently in images from most data sources, and so usually need to be introduced to the dataset by artificial means. This may occur before the source image is scanned by staging, or may be simulated after scanning by threat image projection. Staged threats may use genuine threat objects or whatever facsimiles or surrogates are available. Data can also be entirely synthesised from digital models using X-ray simulation techniques.[20] Benign images may also be staged or synthesised, as well as drawing from routine stream of commerce.

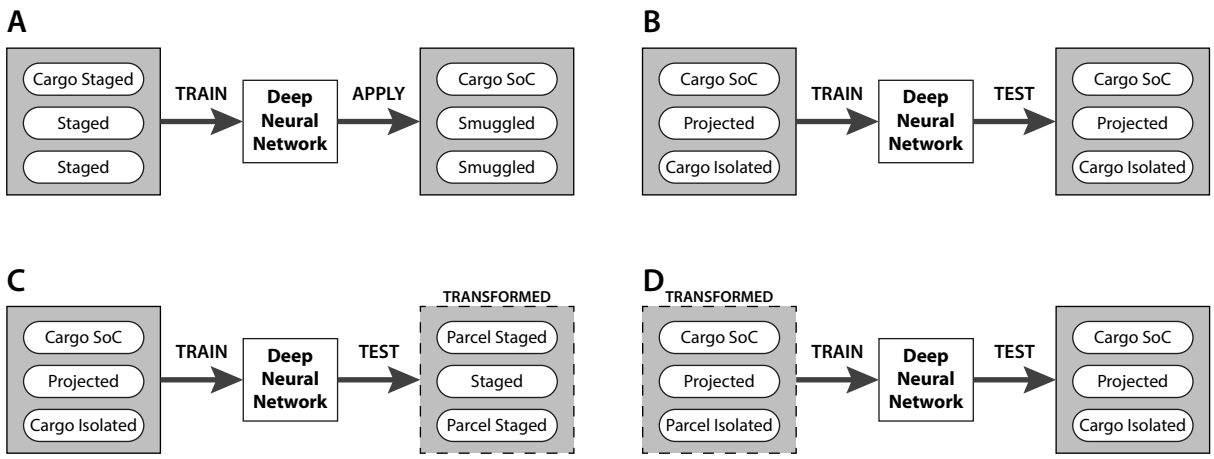

Figure 3. Example configurations for ATD: (**A**) The ideal scenario, where the CNN is trained on a large dataset of images staged to be indistinguishable from SoC, with the hope of perfect CNN transfer/generalisation from training and testing to deployment. (**B**) A configuration we have considered in previous work,[21] where projection is used to insert threats into benign images for training and testing. We consider two approaches to transfer learning. (**C**) *Testing Transfer*: a trained network is tested and applied to data that differs significantly from the dataset it was trained with, potentially requiring conversion of the new data to make it sufficiently similar to what the network expects (see Section 5.1). (**D**) *Training Transfer*: some data different to the intended end target is used during training, but that data is modified to resemble the intended target (see Section 5.2).

A *dataset* for ATD training or testing consists of a set of training/test images. Here we define a dataset as having the same key attributes—background (and benign) image source, threat image source and means of generation—over all contained images. Thus these three properties serve to characterise a dataset. It is essential for supervised learning that the dataset contains examples of all relevant classes, which in the case of ATD means both benign images and images containing threats.

4. APPROACHES TO TRANSFER LEARNING

Transfer learning refers to scenarios where a learned representation or procedure is applied beyond the original training domain. This is distinct from the standard machine learning scenario where a trained model is applied to data that, though previously unseen, is assumed to be drawn from the same underlying population as the training data (Fig. 3A). In transfer learning, there will be some definable difference between the populations.

Using the ATD dataset model described in the previous section, we consider transfer learning to be occurring when one or more of the three dataset properties (background, combination and threat) differs between training and application sets. Because of the low availability of real smuggled threat data, there will usually be some differences.

Fig. 3A details a scenario where the training and testing dataset is staged, with threats physically placed into the containers which are then X-rayed. Ideally the CNN's capabilities would transfer to deployment, where the means of combination change from staged (training/testing) to smuggled (deployment). However there are still opportunities for differences that may hinder this. For example, if a modest range of benign bags are packed with a range of threats in different combinations for training then features of the benign contents could be falsely learned as threat features.

Fig. 3B details a scenario where staging is impractical so the network is trained and tested on a dataset where threats are projected into SoC X-ray images, and it is then hoped that the CNN will generalise to deployment where threats are smuggled. The success of this requires a representative range of input threat data for training and a realistic threat projection.[21]

For cases where components of the training dataset are too limited for the above schemes, we distinguish two general forms of transfer learning according to whether the required data transformations are applied to the training or application dataset.

In the first form, which we term *training transfer* (Fig. 3D), the final application modality is already known and the network is trained with that target in mind. Data is recruited from some other source to provide or augment the training set, but that data is modified to resemble the intended target.

In the second form, *testing transfer* (Fig. 3C), the network is trained on a particular dataset as if that dataset were the target. When the trained network is then transferred to a different dataset, the new set must be adapted to match the parameters of the original training data. Unlike training transfer, testing transfer does not need to be 'premeditated'. Potentially the network can be applied to any modality whose images are not too different from the training modality. In addition to unmodified testing transfer, it is also possible to use a transferred network as a starting point for additional training—*fine tuning*—using new examples drawn from the new domain. This could potentially be useful in situations where there is limited new data for training. The assumption is that an ATD representation will make use of similar features across different data sets and such features will already have been learned in the original context.

5. EXAMPLES: CARGO CONTAINERS VS PARCELS

We illustrate the approaches shown in panels C and D of Fig. 3, discussed in Section 4, with examples using data from cargo containers and parcels.

Cargo data was provided by Rapiscan Systems and consists of approximately 100,000 stream of commerce full-container images, together with 679 individual threats staged in isolation within a cargo container and imaged by the same scanner model (Rapiscan Eagle R60). These threats were extracted manually and used for threat image projection as previously described.[21] Images were obtained as dual energies (4MeV and 6MeV) with a spatial resolution of the order of a few mm and 16-bit precision.

Parcel data was provided by the UK Home Office Centre for Applied Science and Technology (CAST). The data consists of 864 images of staged benign parcels, 2,700 images of staged threat parcels and 1,104 isolated threat images. Images were obtained from two distinct parcel scanners, the models and manufacturers of which are not disclosed. Data were provided as 8-bit per channel RGB images, false coloured using proprietary transfer functions, details unknown. Spatial resolution was also undisclosed, but from objects in the images estimated to be approximately 1.1mm and 0.8mm. X-ray energy was undisclosed but will be lower than that used for container imaging, where the steel skin of the container must be penetrated.

For consistency, we elected to treat all images as if they were single energy only. While the false colours of the parcel images are presumed to encode additional material information related to the channel differences in dual energy data, this information cannot be recovered without knowledge of the transfer functions. We assume that the single energy channel information for these images is approximated by the greyscale value, calculated via a standard weighted sum

$$Y = 0.299R + 0.587G + 0.114B. \tag{1}$$

This assumption may not be reliable and necessarily fails to account for imaging differences due to the unknown energies of the unknown scanners, but it provides a baseline for further adjustments.

Following greyscale conversion, images from the two parcel scanners were blurred to simulate reduced imaging resolution and downsampled to approximately the same scale as the cargo images. Bulk histograms of the three image sets were then calculated (considering the two parcel scanners separately, and excluding the empty backgrounds outside each parcel). Separate transfer functions were calculated for each parcel scanner to shift its histogram closer to that of the cargo scanner. This was inevitably a very rough conversion given the substantial differences in content between the images.

All conversion adjustments were performed in 32-bit floating point arithmetic to avoid data losses due to the 8-bit precision of the original parcel images. Before submission to the CNN, images that had been converted or undergone other modification such as threat projection were scaled and clamped to the 16-bit resolution of the cargo scanner to reduce the likelihood of introducing unrealistic precision cues.

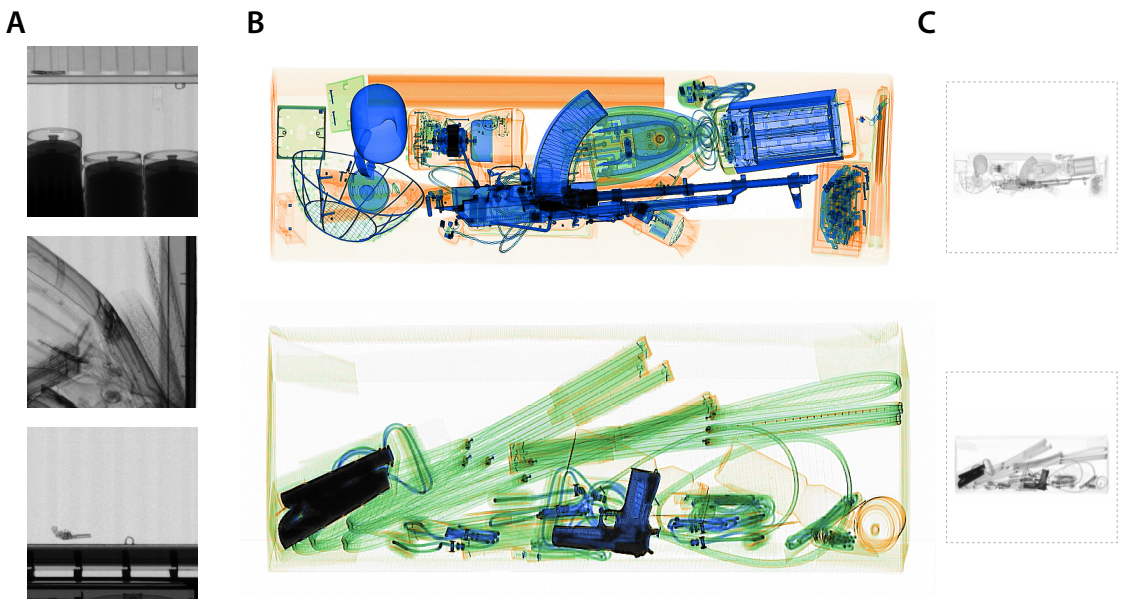

Figure 4. (**A**) Examples of threats projected into SoC container images with which the CNN was trained. (**B**) False colour images of staged threats in parcels, imaged by two different scanners with different resolutions, energies and transfer functions. (**C**) The same images converted to greyscale, downsampled and contrast adjusted for use with the cargo CNN.

The CNN used was largely as described in 1, an adaptation of the well-known VGG-19 architecture[12] newly implemented in TensorFlow. Our network takes as input a 256×256 pixel image patch with two data channels, the raw greyscale values along with an inverted log transformed version that serves to accentuate details in very dark image regions. For large cargo X-rays, we would normally apply the network multiple times to overlapping 256×256 patches, and take the maximum prediction value over the whole container. However, when parcel images are reduced to the same scale as the cargo X-rays, an entire parcel typically fits within a single patch. For consistency in these tests we compared single patches for both modalities, but results for fully tiled container images are included in Table 1 for comparison.

5.1 TESTING TRANSFER: CARGO TO PARCELS

In the first example, we used a CNN that had been trained using SoC container images as background and isolated threats imaged by a container scanner, with the latter added to the former by projection.[21] Some examples of the data used in training are shown in Fig. 4A. The trained network was applied without modification to evaluate staged parcel data. The staged images were converted as described above and, where necessary, padded with plain white background to fill a 256×256 patch. Example parcel images including threats are shown in Fig. 4B, while the same images following conversion are shown in Fig. 4C. It can be seen that the transformations make the images more like cargo data but there are still clear differences.

Evaluation results for all 3,564 staged parcels (864 benign and 2,700 with threats) are shown in Table 1, along with results for 15,000 random cargo patches and 10,000 full cargo container images. It can be seen that the network performs much worse on the parcels than on cargo data, with a false alarm rate of roughly 1 in 2 for 90% detection, compared to 1 in 500 for cargo patches. While this is somewhat better than chance (1 in 1.1 for 90% detection), indicating that *some* transfer of detection capability has occurred, it falls far short of the level required for practical deployment.

A clue to this loss of performance can be seen in the threshold results. For the cargo data, detection thresholds are relatively high—that is, most threats are identified with a high degree of confidence. For the parcel images, however, thresholds are extremely low: to achieve 95% detection, the threshold has to be set at less than one hundredth of one percent indicating that the issue arises from poor detection in the transformed parcel images,

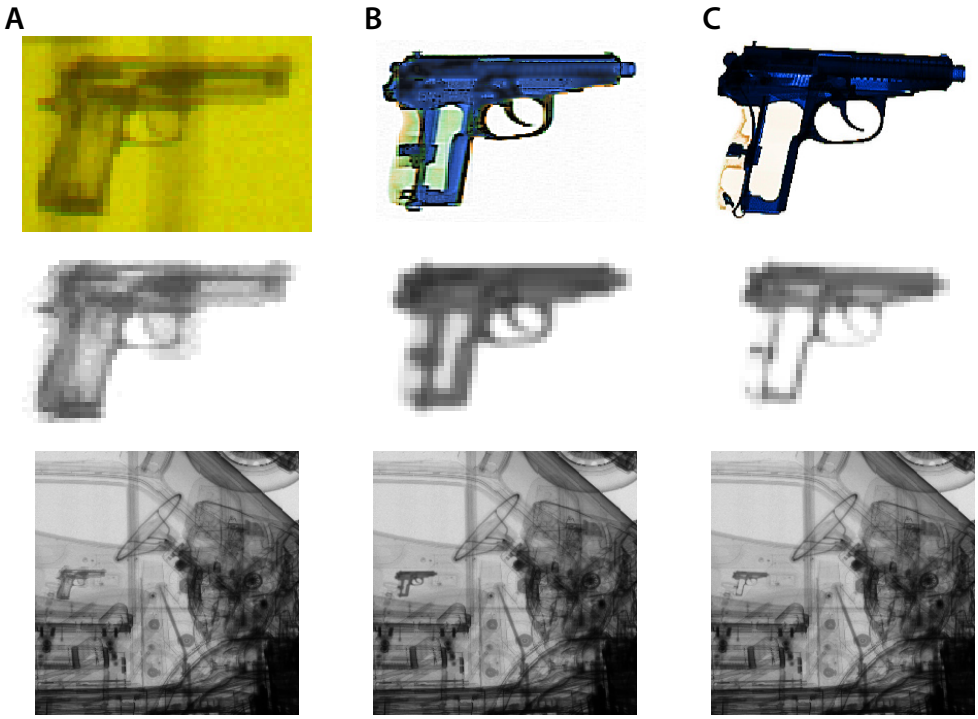

Figure 5. Example threats from: (**A**) the original cargo scanner; (**B**,**C**) two different parcel scanners. Top row shows original images, with obvious differences in resolution and colouring (cargo scan is dual energy). Middle row is the masked, blurred, scaled and contrast-adjusted version prepared for use in TIP. Bottom row shows each image projected into a moderately-cluttered cargo patch.

rather than excessive false alarms. We suspect that the issue is our failure to accurately match the blur of cargo images in the transformed parcel images. Another possibility is that the CNN, when applied to cargo is exploiting projection artefacts which are not present in the staged parcel data. Although the analyses in [21, 22] seemed to rule this out.

Table 1. Evaluating parcel data with a CNN trained on cargo. (Abbreviations: AUC, area under the receiver operating characteristic curve; FPRd, false positive rate at d% detection; Thr$d$, CNN output confidence threshold for d% detection.)

	Evaluated on	AUC (%)	FPR90 (%)	Thr90	FPR95 (%)	Thr95
1	Cargo patches	99.57	0.19	96.06	1.68	55.35
2	Tiled cargo container	98.26	2.04	99.44	11.36	93.21
3	Staged parcel patches	79.85	54.40	0.01	68.17	0.00

5.2 TRAINING TRANSFER: PARCEL WEAPONS TO CARGO

For a second example, we considered the transfer of data from the parcels dataset for use in training of the cargo network. This simulates the scenario where not enough cargo threats are available to adequately span the space of threats, but a good range of threats is available from parcel data. In this case the staged parcels were not used, only the isolated weapon images from the parcel scanners. As before, these were converted to greyscale, blurred, downsampled and adjusted to have a similar intensity profile to the cargo data. Note that the parameters of these adjustments were determined under the limitation that no cargo weapon images were available for comparison, i.e. the transfer functions were estimated from benign images only and were thus necessarily approximate. The isolated weapon images did not have major extraneous details needing removal, but some background noise was

present and this was masked away. These weapons were then used for threat image projection exactly as is done with cargo threats. A comparison of an actual cargo threat image with a similar threat imaged by the two parcel scanners is shown in Fig. 5, both before and after conversion and also as projected into a cargo patch.

The CNN was then trained from scratch using (i) cargo patches with projected parcel threats, and (ii) cargo patches with projection of only 5% of the available cargo threats (28 images). As a further test, the network trained on parcel threats was further 'fine tuned' (i.e., given additional training epochs with new data) using the 5% cargo threat subset, to see if the two datasets could together achieve comparable performance to the full cargo threat set. Each of these trained models was then tested on 15,000 random cargo patches with cargo threat projection. In all cases, both the weapons and the cargo containers used in testing were disjoint from those used for training. Results for these tests are in Table 2.

Table 2. Cargo threat evaluation results for cross-trained networks. (Abbreviations: AUC, area under the receiver operating characteristic curve; FPRd, false positive rate at d% detection; Thr$d$, CNN confidence output threshold for d% detection.)

	Training threats	AUC (%)	FPR90 (%)	Thr90	FPR95 (%)	Thr95
1	All cargo weapons	99.57	0.19	96.06	1.68	55.35
2	5% subset of cargo weapons	99.29	0.55	76.82	2.52	20.85
3	Parcel weapons	92.38	29.20	0.17	43.03	0.05
4	Parcel weapons + 5% cargo fine tune	99.42	0.59	80.69	2.87	30.93

Comparison of rows 1 and 2 show the importance of this scenario—if only a small number of threats are available for training, there is considerable impact on performance as false positive rates are roughly doubled. Row 3 shows that training transfer of parcel threats to cargo produces a system with performance, though much better than with test transfer (Table 1), still below deployable level. Comparison of rows 2 and 3 underlines that we would be better off training with a small dataset of threats and no transfer, than a larger dataset with transfer. Row 4 compared to row 2 shows that we possibly get some slight utility from the transferred training data, when used in combination with the small non-transferred dataset, but the impact is at best very small.

The small utility of the transferred data cannot here be attributed to a mismatch between projection and staging as was a possibility with the section 5.1 experiment. Again it seems that the explanation is most likely a failure to match the blur and contrast characteristics.

6. SUMMARY & CONCLUSIONS

We have made a case for the importance of transfer within the field of Automatic Threat Detection, owing to frequently limited datasets. Through our ontology we have identified three attributes that define an ATD dataset and act as entry points for transfer routines; namely the sources of the image backgrounds and threats, and the means through which they are combined. We have identified imaging parameters that must be considered when transforming datasets for domain transfer, and have demonstrated that distinct types of transfer can be applied at the training and testing phases. Both the approaches of transforming data to comply with a set that is better understood, and cross-training the CNN over a rich and then the sparse domain provide little benefit and are not a substitute for real and extensive domain-specific data. Possibly better transfer can be achieved with better data conversion but detailed information and algorithms used by the scanning devices are proprietary, and without access to this we have found that converting data between scanners is difficult. Human vision is an existence proof that useful data transfer is possible, but we do not know whether this can be achieved in the ATD application simply by improving our transfer methods, or whether it requires different types of network that more readily learn generalisable representations.

ACKNOWLEDGMENTS

Cargo images were provided by Rapiscan Systems while bag and parcel images were provided by the UK Home Office Centre for Applied Science and Technology as part of the Borders X-Ray Image Library.

REFERENCES

[1] Jaccard, N., Rogers, T. W., Morton, E. J., and Griffin, L. D., "Automated detection of smuggled high-risk security threats using Deep Learning," *In: Proc. Int Conf on Imaging for Crime Detection & Prevention* (2016).

[2] Jaccard, N., Rogers, T. W., Morton, E. J., and Griffin, L. D., "Detection of concealed cars in complex cargo X-ray imagery using deep learning," *Journal of X-Ray Science and Technology* (Preprint), 1–17 (2016).

[3] Rogers, T. W., Jaccard, N., and Griffin, L. D., "A deep learning framework for the automated inspection of complex dual-energy X-ray cargo imagery," *In: Proc. SPIE Defense + Commercial Sensing* (2017).

[4] Pan, S. J. and Yang, Q., "A survey on transfer learning," *IEEE Transactions on knowledge and data engineering* **22**(10), 1345–1359 (2010).

[5] Akçay, S., Kundegorski, M. E., Devereux, M., and Breckon, T. P., "Transfer learning using convolutional neural networks for object classification within X-ray baggage security imagery," *In: Proc. IEEE Int Conf on Image Processing* , 1057–1061 (2016).

[6] Cireşan, D. C., Meier, U., and Schmidhuber, J., "Transfer learning for Latin and Chinese characters with deep neural networks," in [*Neural Networks (IJCNN), The 2012 International Joint Conference on*], 1–6, IEEE (2012).

[7] Huynh, B. Q., Li, H., and Giger, M. L., "Digital mammographic tumor classification using transfer learning from deep convolutional neural networks," *Journal of Medical Imaging* **3**(3), 034501–034501 (2016).

[8] Mehta, D., Rhodin, H., Casas, D., Sotnychenko, O., Xu, W., and Theobalt, C., "Monocular 3D human pose estimation using transfer learning and improved CNN supervision," *arXiv preprint arXiv:1611.09813* (2016).

[9] Glorot, X., Bordes, A., and Bengio, Y., "Domain adaptation for large-scale sentiment classification: A deep learning approach," in [*Proceedings of the 28th international conference on machine learning (ICML-11)*], 513–520 (2011).

[10] Daumé III, H., "Frustratingly easy domain adaptation," *arXiv preprint arXiv:0907.1815* (2009).

[11] Griffin, L. D., Wahab, M. H., and Newell, A. J., "Distributional learning of appearance," *PLoS One* **8**(2), e58074 (2013).

[12] Simonyan, K. and Zisserman, A., "Very deep convolutional networks for large-scale image recognition," *arXiv.org* (2014).

[13] He, K., Zhang, X., Ren, S., and Sun, J., ": Surpassing human-level performance on imagenet classification," in [*Proc. ICCV*], 1026–1034 (2015).

[14] Erhan, D., Bengio, Y., Courville, A., and Vincent, P., "Visualizing higher-layer features of a deep network," Tech. Rep. 1341 (2009).

[15] Martonosi, S. E., Ortiz, D. S., and Willis, H. H., "12. evaluating the viability of 100 per cent container inspection at america's ports," *The economic impacts of terrorist attacks* , 218 (2007).

[16] The International Air Transport Association, "IATA Annual Review 2017," (2017).

[17] Royal Mail plc, "Financial report for the full year ended 26 March 2017," (2017).

[18] US Customs and Border Protection, "Container Security Initiative (CSI) Fact Sheet," (2012).

[19] Rogers, T. W., Jaccard, N., Morton, E. J., and Griffin, L. D., "Automated X-ray image analysis for cargo security: Critical review and future promise," *Journal of X-ray science and technology* **25**(1), 33–56 (2017).

[20] White, T. A., Bredt, O. P., Schweppe, J. E., and Runkle, R. C., "Development of a detector model for generation of synthetic radiographs of cargo containers," *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **266**(9), 2079–2089 (2008).

[21] Rogers, T. W., Jaccard, N., Protonotarios, E. D., Ollier, J., Morton, E. J., and Griffin, L. D., "Threat Image Projection (TIP) into X-ray images of cargo containers for training humans and machines," *IEEE International Carnahan Conference on Security Technology* (2016).

[22] Rogers, T. W., *Automated analysis of X-ray images for cargo security*, PhD thesis, UCL, Department of Computer Science (8 2017).