Deep learning for 2D passive source detection in presence of complex cargo

To cite this article before publication: Weston Baines *et al* 2020 *Inverse Problems* in press https://doi.org/10.1088/1361-6420/abb51d

Deep learning for $2D$ passive source detection in presence of complex cargo

26 August 2020

W. Baines[1], P. Kuchment[2], and J. Ragusa[3]

[1] Mathematics Department, Texas A&M University, College Station, TX, USA
[2] Mathematics Department, Texas A&M University, College Station, TX, USA
[3] Nuclear Engineering Department, Texas A&M University, College Station, TX, USA

Abstract. Methods for source detection in high noise environments are important for single-photon emission computed tomography (SPECT) medical imaging and especially crucial for homeland security applications, which is our main interest. In the latter case, one deals with passively detecting the presence of low emission nuclear sources with significant background noise (with Signal To Noise Ratio (SNR) 1% or less). In passive emission problems, direction sensitive detectors are needed, to match the dimensionalities of the image and the data. Collimation, used for that purpose in standard Anger γ-cameras, is not an option. Instead, Compton γ-cameras (and their analogs for other types of radiation) can be utilized. Backprojection methods suggested before by two of the authors and their collaborators enable detection in the presence of a random uniform background. In most practical applications, however, cargo packing in shipping containers and trucks creates regions of strong absorption and scattering, while leaving some streaming gaps open. In such cases backprojection methods prove ineffective and lose their detection ability. Nonetheless, visual perception of the backprojection pictures suggested that some indications of presence of a source might still be in the data. To learn such features (if they do exist), a deep neural network approach is implemented in 2D, which indeed exhibits higher sensitivity and specificity than the backprojection techniques in a low scattering case and works well when presence of complex cargo makes backprojection fail completely.

Keywords: source detection, Compton camera, illicit nuclear material

Submitted to: *Inverse Problems*

1. Introduction

Checking for presence of illicit nuclear materials (most probably in small quantities and shielded by cargo) at border crossings and shipping cargo containers in harbors is an important homeland security task. Ideally, one would try to reconstruct from the detected signals the source distribution inside the cargo. When the data is sufficiently well behaved (e.g., in SPECT), analytic reconstruction is often possible [30]. However, in a very low SNR environment, as in the case of illicit nuclear source detection, this is impossible. Indeed, the forward analytic (integral transform type) models are not applicable. Moreover, even if they were, attempts of any filtration in FBP-type techniques lead to reconstruction deterioration. The saving grace is that in this case practitioners are mostly interested in getting reliable (i.e., with low rates of false positives and false negatives) information about the presence of a source, rather than its exact location.

In passive emission imaging, detectors must be direction sensitive. Indeed, otherwise the data measured has insufficient dimension for recovery of an image. Directional information is especially critical when SNR is too low for the intensity fluctuations that arise due to the presence of a source to be statistically significant. The following options for obtaining directional sensitivity are available:

- *Mechanical collimation*, when only rays incident along (or close to) a certain line are allowed to reach the detector (see Section 2). This, while determining the incoming photon's direction, significantly reduces the signal strength and thus becomes unsuitable for low SNR.
- *Compton γ-cameras* represent a more recent, and gaining its appreciation, type of γ radiation detectors that determine a surface cone of possible incident trajectories, rather than the exact directions.
- Neutron detectors are being developed that (albeit based on different physics principles) produce similar cone information and lead to similar mathematical analysis.

Backprojection detection technique introduced in [5, 33] relied upon finding suspicious **locations**. It utilized the following three assumptions:

(i) geometric smallness of the source (usually of linear dimension on the order of 1% of the linear cargo size);

(ii) existence of a sufficient number of particles from the source reaching the detector being **ballistic** (non-scattered);

(iii) unstructured strong random background.

The idea is rather simple: backprojecting the incoming trajectories (or, in the Compton case, the whole surface cones of possible trajectories) of particles, one hopes that maybe, due to sufficient presence of ballistic particles detected from the source, one can see a statistically significant accumulation at the geometrically small source's location (see Fig. 1)

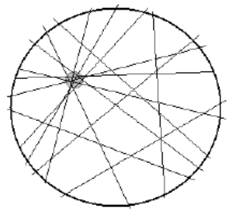

Figure 1. An idea of the backprojection method.

Analysis done in [5] provided a crude formula for the total number N of particles (and thus observation time) needed to make detection with high (on the order of 99%) sensitivity and specificity (i.e., with low levels of false negatives and false positives).

$$N \gtrsim \left(\frac{8}{S}\right) p(1-p). \qquad (1)$$

Here p is the ratio of the linear dimension of the source relative to the dimension of the cargo and S is the SNR, defined as the proportion of the ballistic particles from the source versus the total number of source and background particles. In the cases considered in [5] N had to be on the order of 600000, which is not unrealistic for γ photons **not** screened by heavily shielding cargo. High specificity has been hardwired into the method, so satisfying (1) was only needed in [5, 6] to ensure high sensitivity.

The implementation of the technique worked as follows [5, 6]: the data was backprojected, which resulted in a large background level throughout the volume. When the object was completely surrounded by detectors, this level was essentially constant and the mean was removed. When the detectors did not surround the object completely (e.g., no detector below the object), the global mean is irrelevant, and at each location the mean over a smaller patch was removed. After this clean-up the locations with an intensity less than five standard deviations above the mean suggested by the Central Limit Theorem were cut off. The results were interpreted as indications of a source being present. Thousands of Monte Carlo simulations showed that the inequality (1) performs well and if N is at or above this threshold, detection occurs with high sensitivity and specificity ‡.

This technique works reasonably well in the absence of complex cargo, but starts failing if such cargo is present [6], due to the second and third assumptions being inapplicable§. However, visual inspection of the backprojected data (see [6]) seems to indicate that the data **might** still contain a signature of the source presence. Indeed, when the method of [5] was applied to some cases of complex cargo in [6], despite its failure to detect presence of the source, such signatures (e.g., different highlighting of the pathways between cargo boxes) seemed to appear only when a source was present (see Figure 2). The reader should

‡ An alternative Bayesian approach was implemented in [33].

§ This cargo problem is mostly non-existent when detecting neutrons coming from the source. However, some other (non-mathematical) issues arise, such as for instance lower number of particles detected.

take into account that the color scales are different in the three pictures there and assigned automatically by the visualization software. This is of no importance, since it is not the intensity, but rather the patterns of highlighted pathways between boxes seem different.

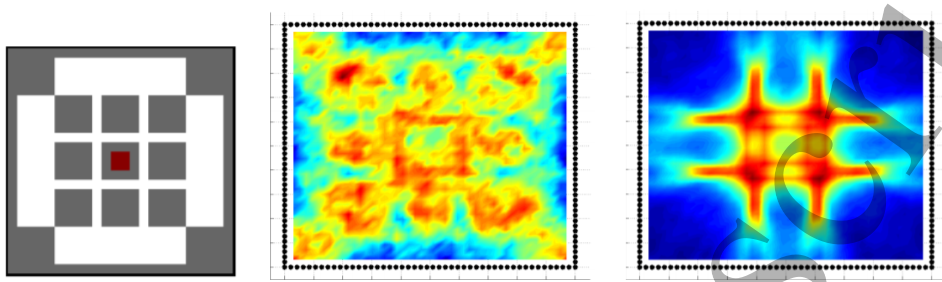

Figure 2. (Left): Example of complex cargo configuration for which backprojection methods fail (i.e., no statistically suspicious locations are found). The red spot denotes the source location, the grey area represents iron and the white area represents air. (Middle): Backprojection results in absence of source. (Right): Backprojection results in presence of source.

No model of this effect has been developed, no telling features have been learned, and thus no detection algorithm came out of such observations.

This has led the authors to attempt deep learning for the source inference in the hope that a network could learn what we could not. Our main goal is to detect the presence/absence of a source, not necessarily its location. If there is high probability of presence of the source, in practice one would check the cargo with other (hand-held) devices. However, one also needs to achieve high specificity, to avoid large numbers of false positives.

One should note that quality of tomographic image reconstructions using neural networks has been questioned recently, see e.g. [7,8]. This critique, however, does not apply to the problem at hand, where we only look for a binary output rather than an image.

We describe now the structure of the article. Section 2 contains a brief description of the Compton type cameras and references to the known analytic approaches. Success of deploying neural networks is predicated upon our access to sufficient data for neural network training. Thus, the first step - generating various complex cargo scenarios is described in Section 3. To avoid the inverse crime (overfitting), different processes of generating cargos are used for creating training and testing samples. Then, in absence of real data (which would require having weapons grade nuclear materials and physically creating thousands of different cargoes), we use (Section 4) the technique of forward radiation transport simulations customarily used in nuclear engineering. As has been mentioned, the actual type of radiation is mathematically irrelevant, but to be close to real world scenarios and numerical parameter values, the case of γ-photons coming from an U-238 source and real world material parameters for cargo are used. The design of the network is described in section 5. The results are presented in Section 6. Additional remarks can be found in section 7. Acknowledgements are provided in section 8. The algorithm description is located in the Appendix.

2. Collimated and Compton γ-Cameras

Mechanical collimators (see Figure 3) can be installed in front of a direction insensitive $\gamma-$camera to block all particles but those incident along (or close to) a desired trajectory.

Mechanical collimators are widely used in medical imaging. They, however, significantly

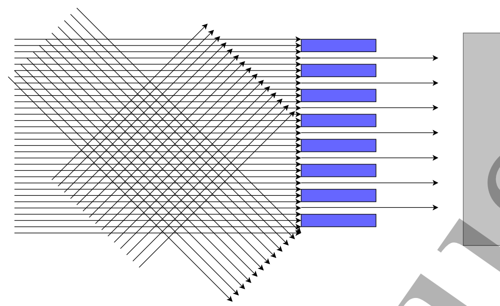

Figure 3. Light collimation diagram

attenuate the signal and require rotating the detector (or the object). In the applications with sufficiently high SNR, this additional data loss is not such a problem. In dealing with low SNR signals however, this renders recovery of weak signals impossible. For this reason one can consider Compton type cameras instead.

The **Compton camera** is a type of γ-particle detector‖ that does not attenuate the incident particles. The price to pay is that it provides less precise direction information than collimation would give. Namely, only a surface cone of possible incoming directions is measured rather than a precise trajectory (see Fig 4). In the absence of mechanical

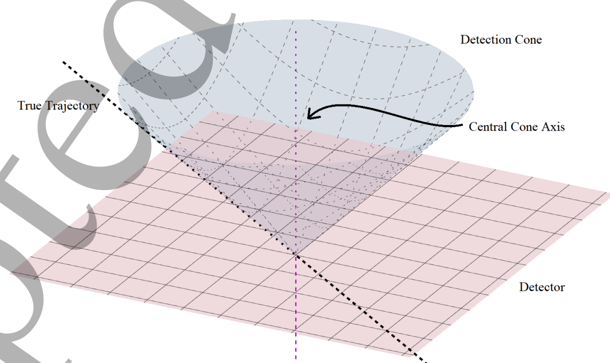

Figure 4. Surface cone produced by Compton camera from particle detection

collimation, signal strength is effectively maintained, although the directional information is less precise and thus data analysis becomes more complex. On the other hand, the data provided is significantly over-determined (e.g., the space of cones in $3D$ is five-dimensional,

‖ As we have mentioned before, novel neutron detectors (albeit based upon different physics rather than Compton scattering) that provide mostly similar cone information are currently being developed.

versus the unknown distribution being three-dimensional). This turns out not to be a bad thing at all, but rather a blessing for stable inversion (see [30] for details and further references).

A variety of exact inversion formulas from Compton data of filtered-backprojection and other types have been developed and implemented (see [30] and references therein). The choices are much more diverse than for the usual Radon transform inversions (see [24]). The reason is that the Compton data is highly overdetermined. It was shown that this feature can be used to get high quality reconstructions in SPECT in presence of 50% noise and higher. However, this is a far cry from the low SNRs encountered in the homeland security problems described above.

3. Simulating Cargo Scenarios

If one intends to tackle a problem using deep learning, it is natural to start by acquiring large amounts of training and testing data.

In order to obtain rich training data for a neural network, at least thousands (better hundreds of thousands or millions) of cargo configurations are needed. Due to the sensitive nature of the materials involved in this work, we are unable to procure real-world data, so we resort to synthetic simulation. The high computation costs of these simulations restricted us to several thousands of samples, reaching up to 4×10^5. However, our results (see Section 6) already show a success in detection.

To start, we randomly produce several thousand cargo configurations and compute forward radiation data simulations with up to four randomly placed sources and without them for each one. In order to avoid overfitting (and an inverse crime), different cargo generation procedures are used for producing the training and testing data.

3.1. Procedural Generation of Training Cargo Configurations

A square cargo hold of size of $2.4m \times 2.4m$ is assumed and partitioned into $2.4cm \times 2.4cm$ cells (the possible source would occupy one of them). Each cell can be indexed via a pair of row and column indices, (i, j), with $1 \leq i, j \leq 100$ and is assigned a material identification number $ID_{i,j}$. These numbers correspond to a variety of materials, including Air, concrete, highly enriched uranium, iron, cotton, wood, plastic, and fertilized (their detailed chemical content described in [6]).

Real cargo typically consists of several boxes with small spaces in between. In order to emulate this, an algorithm is implemented to generate different cargo configurations. It consists of three main steps:

- A network of several horizontal and vertical "corridors" between boxes with random widths and locations is generated. The number of corridors c is selected randomly in a desired range $c_{min} \leq c \leq c_{max}$.

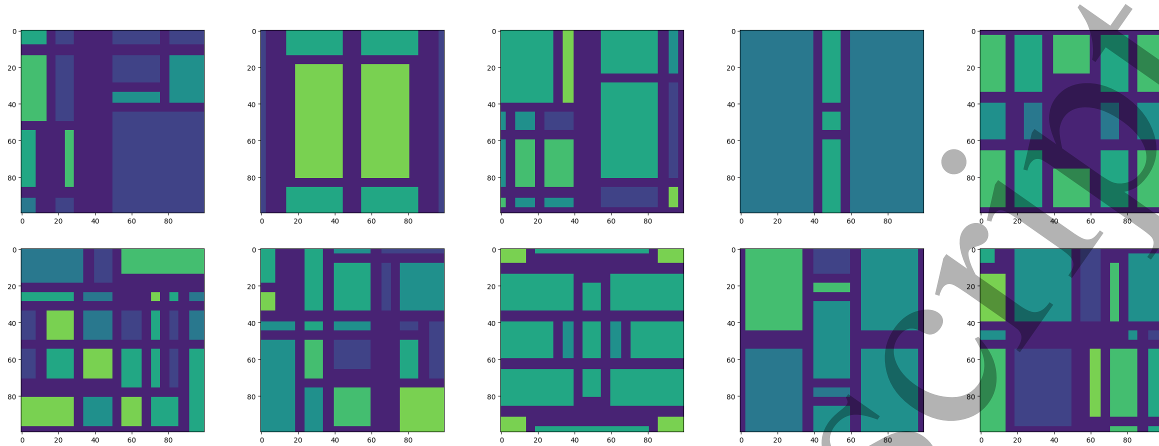

Figure 5. A selection of cargo configurations procedurally generated via Algorithm 1

- The resulting configurations are unlikely to be symmetric, while real cargo might happen to be symmetric. To check whether symmetry plays any role in detectability¶, a portion of the samples produced are "symmetrized" by enforcing various (rotation and mirror) symmetry rules.
- Connected components of the rest of the space are identified as distinct "cargo boxes." Then material contents are assigned to all boxes. In a subset of (rather than all) symmetric cargo configurations, material contents are also "symmetrized" according to the corresponding rule.

Generating the corridors between pieces of cargo is performed using a modification of the procedure outlined in [11] for generating road networks. For the training set, we only use networks consisting of horizontal and vertical segments, while for the testing set tilted and non-orthogonal pathways are allowed.

Remark. Instead of selecting corridor locations uniformly randomly, their locations for training are selected according to a probability distribution generated from a type of gradient noise developed in [25] in order to automate the production of realistic looking textures in computer graphics. A different algorithm is used for testing samples.

Identification of connected components ("boxes") is performed using SciPy's (Scientific Python, a popular Python package for scientific computing [17]) implementation of the algorithms outlined in [32].

The entire generation procedure is summarized in Algorithm 1 in the Appendix (Section 9).

3.2. Procedural Generation of Testing Cargo Configurations

To avoid the inverse crime of overfitting, testing configurations are produced by a somewhat similar, but independent algorithm. Namely, the middle points, the lengths and width of the

¶ Disclosure: Our results show that symmetries do not influence detectability.

corridors are selected randomly and independently. Moreover, the corridors are not required to be vertical or horizontal, or even orthogonal at their intersections anymore. Finding the boxes (connected components of the complement) and filling them with materials is also done randomly, similarly to the training case.

3.3. Source placement

A source of a (randomized) strength corresponding to approximately 1% SNR is placed randomly into the cargo.

Multiple sources $(0, 1, 2, 3,$ or $4)$ are also modeled to see the effect on detection. Two scenarios are used:

(i) when all the sources have the same strength $\approx 1\%$ SNR and

(ii) when the strength of the source is diluted between several locations.

One naturally expects deterioration of the detection in the 2nd case, while *a priori* it would not be surprising if it happened in the 1st as well (although our results will demonstrate that this does not happen). Indeed, the backprojection detection, as well most probably the one by deep networks, if successful, should use some geometric assumptions (e.g., geometric smallness of the source), since the source's strength alone would not be statistically significant. Thus, multiplying the number of sources in principle might degrade the geometric features of importance (albeit one does not know what these are).

4. Forward radiation simulations

As previously mentioned, the nature of particles is irrelevant, but in order to be in realistic situations, the γ particle detection is considered, where the material parameters and emission and background rates that are used assume realistic values.

After the cargo scenario has been created, one needs to simulate training and testing data by solving a massive forward radiation transport computation. Fortunately, reliable simulation tools have been developed by nuclear engineering researchers.

4.1. Physics Preliminary

U-238 (Uranium-238) photons from the 1.001 MeV emission line have mean-free-path in high-Z materials sufficiently high to be detected outside the container (13.3mm mean-free-paths) [28]. In our application, sources of background radiation include a concrete base located some distance below the container. (Cosmic rays and other natural sources can be easily included and do not influence the results much.) These background sources radiate at much higher energies than 1.001 MeV, including 1.461 MeV from Potassium-40, 1.12 MeV and 1.76 MeV from Bismuth-214, and 2.61 MeV from Thallium-208 (Bismuth and Thallium are products of the decay of Uranium-238 and Thorium 232 respectively, and are present in trace amounts in concrete). Gamma photons which downscatter from these sources into the energy group surrounding the 1.001 MeV line account for the noise in our signal. Gamma photons from the source will also undergo scattering and absorption within the volume of the container, which will reduce the number of ballistic source particles reaching the detectors placed around the container, thus weakening the signal.

4.2. Mathematics of the forward radiation data simulation

The radiation transport within the cargo container is modeled by the linear Boltzmann equation, given below using the multigroup approximation:

$$\vec{\Omega} \cdot \vec{\nabla} + \Sigma_t^g(\vec{r})\Psi^g(\vec{r},\vec{\Omega}) = \sum_{g'=1}^{G}\sum_{l=0}^{L}\Sigma_{s,l}^{g'\to g}(\vec{r})\sum_{m=-l}^{l}\Phi_{l,m}^{g'}(\vec{r}) + Q^g(\vec{r},\vec{\Omega}) \tag{2}$$

where $\vec{r} \in \mathcal{D}$ is the position, $\vec{\Omega} \in \mathbb{S}^2$ the set of discrete directions and $g \in [1, G]$ the energy group. $\mathcal{D}$ is the volume of the cargo container, $\mathbb{S}^2$ is the unit sphere, G is the total number of energy groups, Ψ^g is the photon angular flux in the energy group g, Σ_t^g is the total interaction cross section in group g, $\Sigma_{s,l}^{g'\to g}$ is the $l^{th}-$Legendre moment of the scattering cross section from group g' to group g, L is the maximum anisotropy expansion order, and Q^g is the volumetric source of photons in group g (stemming from the U-238 source). The moments of the angular flux are given by

$$\Phi_{l,m}^g(\vec{r}) = \int_{4\pi} Y_{l,m}(\vec{\Omega})\Psi^g(\vec{r},\vec{\Omega})d\Omega \tag{3}$$

where $Y_{l,m}$ is the spherical harmonic of order of l and degree m. Eq. (2) is supplied with boundary conditions:

$$\Psi^g(\vec{r},\vec{\Omega}) = h^g(\vec{r},\vec{\Omega}) \qquad \forall \vec{r} \in \partial\mathcal{D}^- \tag{4}$$

where $\partial\mathcal{D}^-$ is the incoming boundary defined as $\partial\mathcal{D}^- = \{\vec{r} \in \partial\mathcal{D} \text{ such that } \vec{\Omega}\cdot\vec{n}(\vec{r}) < 0\}$ with $\vec{n}(\vec{r})$ the outward unit normal vector at position $\vec{r}$. The function h^g describes the background radiation due to a large concrete slab underneath the container, as previously described. Cross sections for various materials were generated using NJOY-99 [23]. The multigroup structure employed ranges from 1.00099 MeV to 2.61449 MeV with narrow bands centered at the radiation lines of the background and U-238.

For the purposes of this paper, calculations are carried out in two-dimensional space and only the energy group corresponding to the 1.001 MeV line is considered after solving Eq. (2). The photon transport equation, Eq. (2), is discretized using standard techniques:

(i) S_n product Gauss-Legendre-Tchebychev angular quadrature [27] is employed (only a small number of polar angles are needed, but a very high number of azimuthal angles are needed to resolve properly the angular distribution in the 2D domain.)

(ii) Spatial discretization based on a standard bilinear discontinuous finite element technique with upwinding at cell interfaces. [26, 31]

(iii) Transport sweeps and Source Iteration are employed to solve the resulting system. [21]

Once the transport equation (2) has been solved, the outgoing angular photon flux at any boundary edge in 2D is recorded, which serves as the input data for use in Deep Learning and Backprojection.

Once configurations have been generated, a radiating source emitting an expected 8042.17 photons per second at 1.001 MeV is randomly placed, a forward radiative transfer equation is solved, and from its solution the radiation angular flux distribution on the boundary of the cargo is collected.

Due to linearity of (2), the situations of presence of zero to four randomly placed sources could (and were) easily incorporated.

5. Convolutional Neural Network

Using a fully connected network for the problem seems to be hardly feasible even in 2D, less so in 3D, in particular due to high dimensionality of the Compton camera data. The saving grace here is that, as in many imaging problems [22], one expects that the important correlations occur mostly between close pixels, and hence convolutional neural networks, which are much more compact due to weight sharing, offer a hope. We thus construct, train, and test a deep convolutional neural network (CNN). This hand-waving argument for using CNN needs to be confirmed by computations, which is done in this text.

The suggested CNN architecture is summarized in Figure 6 below. The input data dimension is $144 \times 10^3 = 400 \times 360 \times 1$, as we model 400 equally spaced detectors with 360 equally spaced angular bins and only one energy bin is used. The network is trained on 1689 unique simulated cargo configurations with varying numbers of sources present. By exploiting the fact that the Boltzmann equation (2) is linear, we can produce multiple new samples from each configuration by taking varying combinations of sources and detectors. We simulate up to four sources per configuration, and four linear arrays of detectors along each edge of the cargo. This leads to a total of $1689 \times 15 \times 16 = 405360$ total samples. The various combinations are summarized in Table 1 Below. The output of the CNN is two

Number of Sources	One Detector	Two Adjacent Detectors	Two Opposite Detectors	Three Detectors	Four Detectors	Total
0	6756	6756	3378	6756	1689	25335
1	27024	27024	13512	27024	6756	101340
2	40536	40536	20268	40536	10134	152010
3	27024	27024	13512	27024	6756	101340
4	6756	6756	3378	6756	1689	25335
Total	108096	108096	54048	108096	27024	405360

Table 1. Number of training samples in each category

probability measures: $\mathbb{P}_d$ on $\{0, 1\}$ and $\mathbb{P}_n$ on $\{0, 1, 2, 3, 4\}$. A source is determined to be present if $\mathbb{P}(x = 1) > 0.5$, and absent otherwise. $\mathbb{P}_n$ predicts the number of sources present, which we set to $k = \text{argmax}_{0 \leq j \leq 4} \mathbb{P}_n(x = j)$. The loss function used for training is the binary cross-entropy loss:

$$\mathcal{L}(y, \hat{y}) = -y \log \hat{y} - (1 - y) \log(1 - \hat{y}), \tag{5}$$

where y is the network prediction and $\hat{y}$ is the target value (see [15]). The CNN was trained on simulations of a localized source in the presence of high background noise ($SNR = 0.01$). In all cases, early stopping is used to halt training before over-fitting. The various hyper-parameter values used in training are summarized in Table 2 below. The CNN is implemented using Keras with Tensorflow as its backend. Keras is a high level API (Application Programming Interface) for interfacing with machine learning toolkits such as Tensorflow, Theano, and Microsoft Cognitive Toolkit. It helps streamline the construction and training of neural networks [12]. Tensorflow is Google's machine learning toolkit and was chosen due to its scalability, wide range of features, and the wide range of documentation and tutorials available [1]. Any parameters not explicitly mentioned here were set to default values.

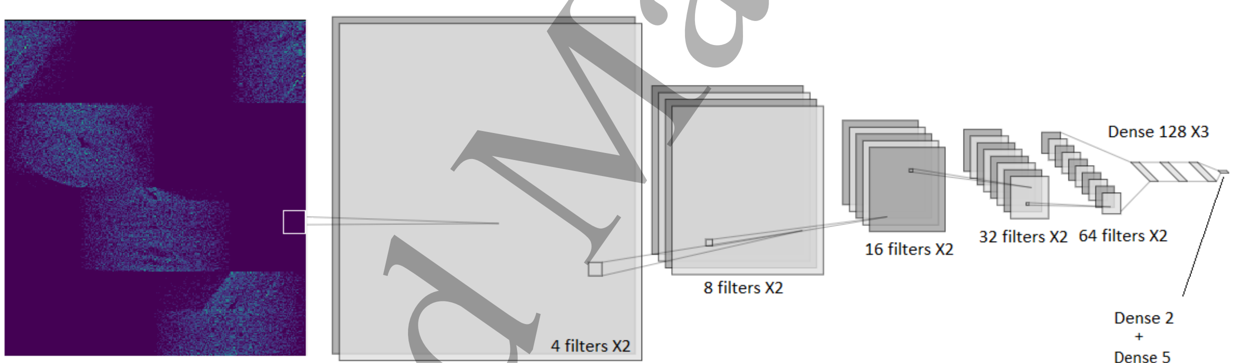

Figure 6. CNN architecture used for source detection. The left-most cell shows an example of the detector data input to the CNN. 2×2 Max pooling layers are placed after every second convolutional layer.

6. Results

After training the CNN, we considered a large variety of cargo scenarios to test and to compare and contrast the performance of the CNN against the backprojection method of [5, 6]. We detail some interesting specific example scenarios in Sections 6.1 and 6.2. We then investigate the statistical performance of the CNN on large scale data sets to evaluate the sensitivity and specificity of the CNN in Section 6.3, and to assess its performance with different numbers of sources and detectors in Section 6.4. Finally, in Section 6.5 we discuss the relation between cargo configuration and exposure time and how this affects the practicality of our technique.

Optimization Method	Adam (See [18])
Activation	RELU (Softmax at output)
Bias	True
Convolution Window Size	3x3
Learning Rate	2.0×10^{-5}
Learning Rate Decay Rate	0
Batch Size	4
Early Stopping Patience	3 epochs
Loss	Binary Cross-Entropy

Table 2. Hyper-parameters used during training

6.1. Example Scenarios

We describe now several (out of many, see later on in this text) sample results of testing the trained network on various scenarios not included in the training set.

6.1.1. Example #1

This configuration (right) as well as backprojection (left) with source present is shown in Figure 7 below. The backprojection procedure described in the Introduction did not lead to any statistically significant detection. We, however, show the raw (not cleaned up) backprojection picture for the reader to notice the corridor highlighting phenomenon observed in [6]. The network, on the other hand, succeeds in detecting presence of the source. This is one of the heavy iron configurations which has a shorter exposure time (18 seconds for 101,180 background particles).

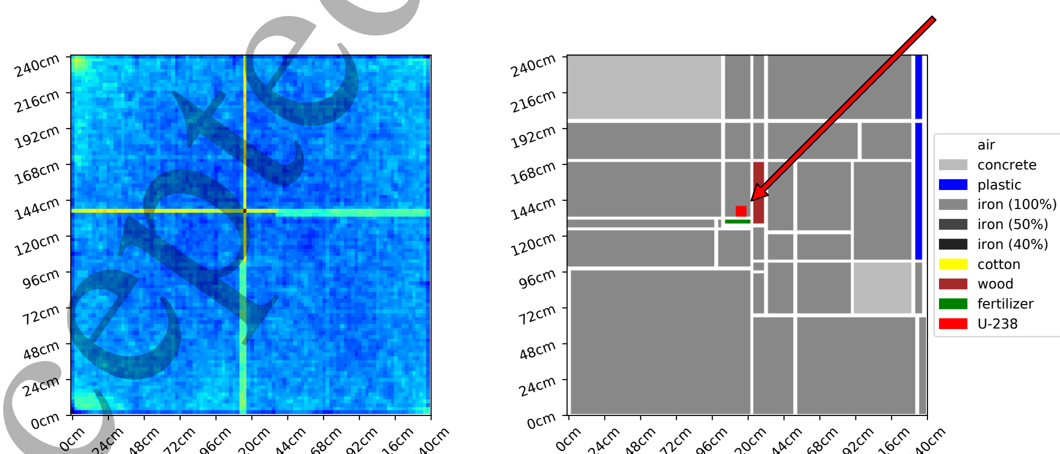

Figure 7. Left: Backprojection with no source detected. Right: Cargo configuration with source location indicated by arrow. 101,180 particles detected, 100,185 background particles and 995 source particles. Exposure time is 18 seconds.

6.1.2. Example #2

Next we consider the scenario shown in Figure 8, where backprojection fails to detect the source (and thus is not shown), but the network succeeds. Here the exposure time needed for the detection is significantly longer. In this configuration a long thick iron slab effectively blocks one side of the detectors. Smaller chunks of iron spread throughout the container further attenuate the signal along certain trajectories. As a result, it would take 9 hours and 26 minutes to detect the needed 101,092 particles. Unless one is talking about a shipping container, this is practically unfeasible. As the results in Section 6.3 show, twice shorter time would still do decently, and even five times shorter time might sometimes be used, although at the expense of higher false positive rate.

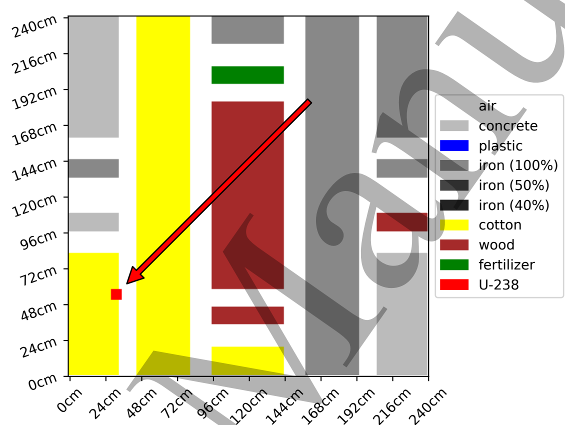

Figure 8. Cargo configuration with source location indicated by arrow. 101,092 particles detected, 100,095 background particles and 997 source particles. Exposure time is 9 hours and 26 minutes.

6.1.3. Example #3

Here we consider a somewhat more tenable scenario shown in Figure 9, where backprojection fails to detect the source, yet the network succeeds. In this case the exposure time is 50 minutes and 17 seconds for 100,866 particles. In this configuration several large blocks of iron are periodically tiled in the container, with the source located within one of the blocks.

6.1.4. Example #4

Now we consider a somewhat extreme scenario (Figure 10), where both approaches succeed in detecting the source. In this case the exposure time is 3 days and 12 hours for collecting 101,272 particles. In this configuration one very large block of iron in the center of the container surrounds the source. The source is still localized relatively well by backprojection for this scenario. Most of the cargo is filled with a homogeneous material, which might explain why backprojection did not fail.

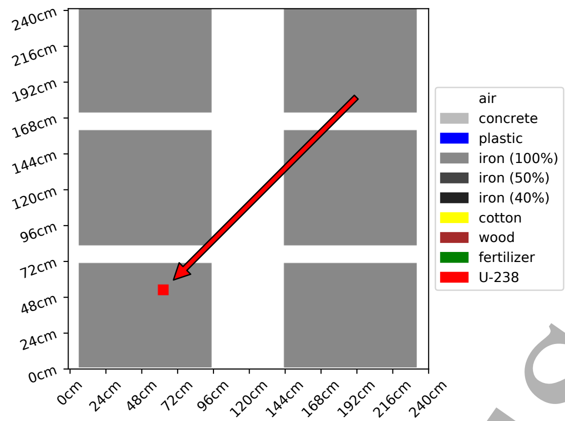

Figure 9. Cargo configuration with source location indicated by arrow. 100,866 particles detected, 99,867 background particles and 999 source particles. Exposure time is 50 minutes and 17 seconds.

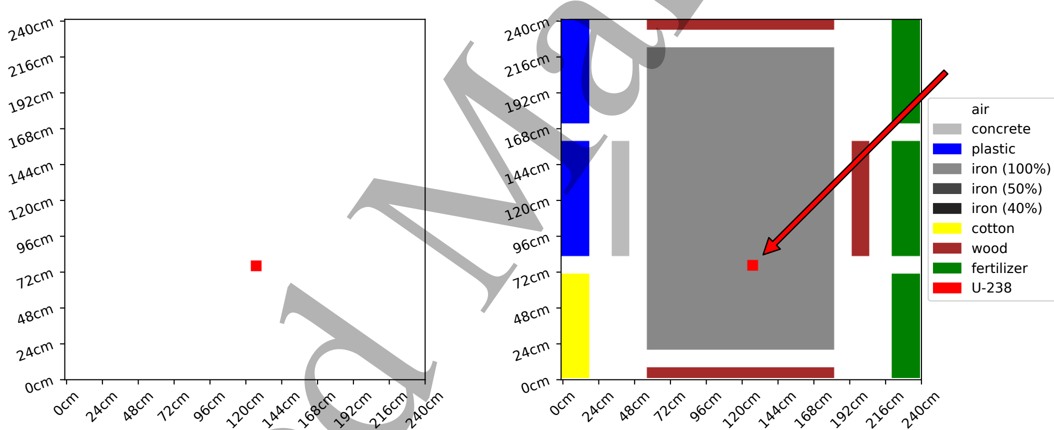

Figure 10. Left: Backprojection with source detected. Right: Cargo configuration with source location indicated by arrow. 101,272 particles detected, 100,328 background particles and 944 source particles. Exposure time is 3 days and 12 hours.

6.1.5. Example #5

Next, we consider a rather easy scenario (Figure 11), where both backprojection and the network succeed. In this case the exposure time is 276 milliseconds for 100,898 particles. In this configuration several small blocks of different materials are spread throughout the container. Only an insignificant amount of particles are scattered, so backprojection recovers the source distribution extremely well.

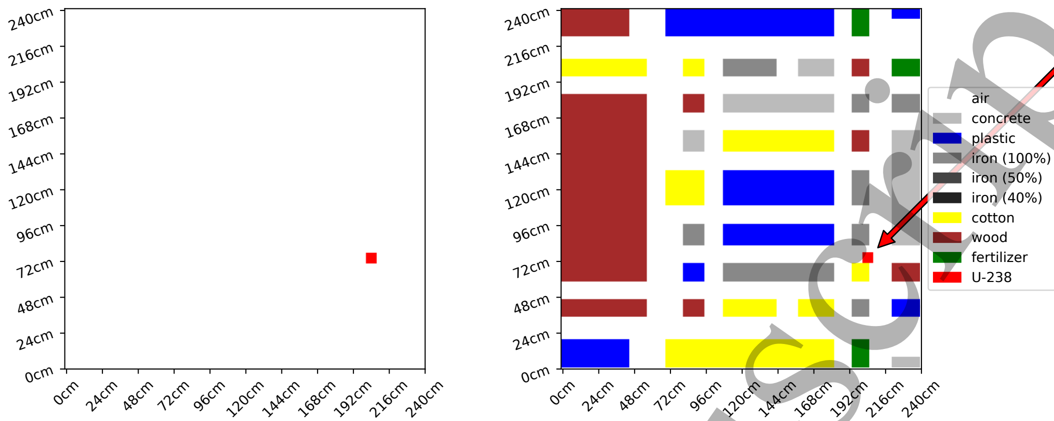

Figure 11. Left: Backprojection with source detected. Right: Cargo configuration with source configuration indicated by arrow. 100,898 particles detected, 99,911 background particles and 987 source particles. Exposure time is 276milliseconds.

6.2. Generalization to more complex scenarios

Now we will include more complex situations, considerably different from the ones used for training. Namely, the corridors are not necessarily aligned vertically and horizontally, nor are intersecting corridors orthogonal. The algorithm of producing configurations was different from the one used in training. Additionally, we allow multiple sources to be present. The results show that the network passes well this generalization test.

6.2.1. Example #6

In this configuration (see Figure 12) several iron blocks are spread throughout the container, but a sufficient amount of low attenuating paths exist between the sources and detector arrays for backprojection to recover the sources well. There are two sources present very near to each other. This clearly aids the backprojection method in successfully detecting the sources. The CNN also succeeds in detecting presence of both of the sources. Here the exposure time is 649 milliseconds for 101,497 particles.

6.2.2. Example #7

In this configuration (Figure 13) several heavy iron blocks cut diagonally through the container slightly off-center. Three sources are present in this scenario, the two sources around the middle are localized well with backprojection, since most of the materials only weakly attenuate the signal, but the source on the other side of the heavy iron has several attenuating materials to contend with, so the backprojection smears its signature throughout the diagonal corridor it's in. Both backprojection and the CNN successfully predict that there is a source, although backprojection fails to locate the third source. This third source may prove difficult for the CNN to contend with as well, as the CNN predicts there are only two

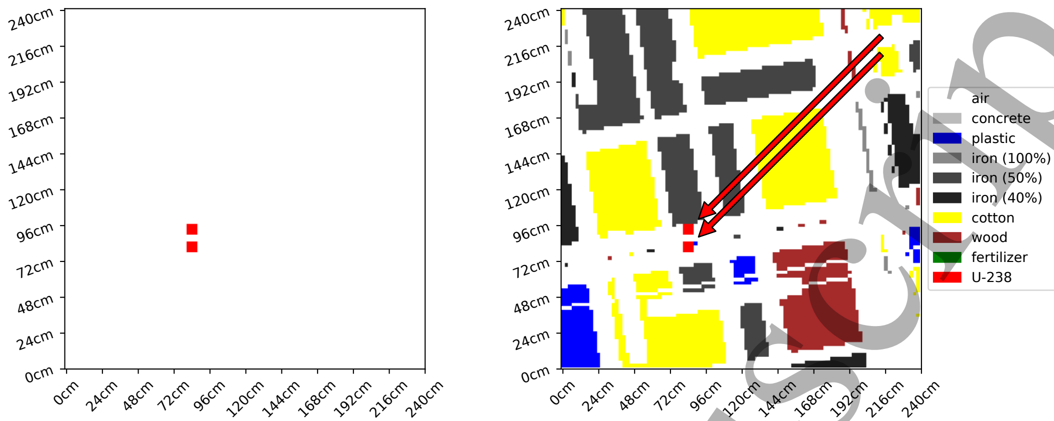

Figure 12. Left: Backprojection with source detected. Right: Cargo configuration with source configuration indicated by arrow. 101,497 particles detected, 99,492 background particles and 2,005 source particles. Exposure time is 649 milliseconds.

sources present. Here the exposure time is 371 milliseconds for 102,790 particles.

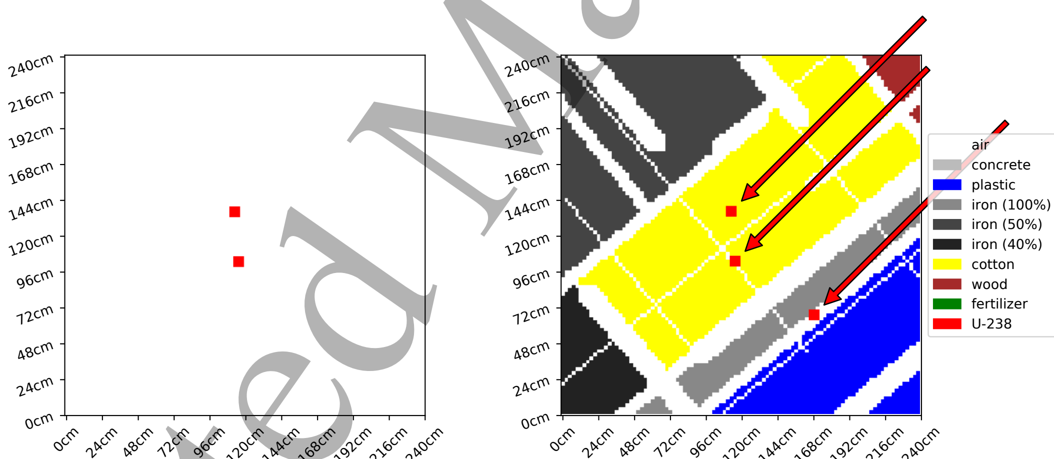

Figure 13. Left: Backprojection with source detected. Right: Cargo configuration with source configuration indicated by arrow. 102,790 particles detected, 99,846 background particles and 2,944 source particles. Exposure time is 371 milliseconds.

6.2.3. Example #8

Here several iron blocks surround the center of the container (Figure 14). Four sources are present in this scenario, two of them directly adjacent (and thus hard to distinguish in the picture) and all four are near the center of the container. In this case backprojection fails to localize any of the sources due the limited angular information in the signal as a result of the attenuating properties of the iron. The CNN, on the other hand succeeds in detecting the presence of all four of the sources, even despite the close proximity of two of them. Here the exposure time is 1.97 seconds for 103,789 particles.

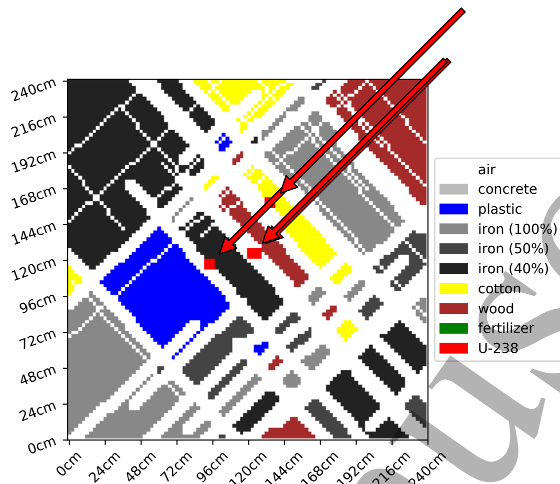

Figure 14. Cargo configuration with source configuration indicated by arrow. 103,789 particles detected, 99,900 background particles and 3,889 source particles. Exposure time is 1.97 seconds.

6.2.4. Example #9

Finally, we consider a simple case where backprojection and the CNN both succeed. In this case there is ample angular information for backprojection to localize the source well and the CNN correctly predicts the presence of a single source. The exposure time is 21.92 seconds for 100,672 particles. The configuration can be seen in Figure 15 below.

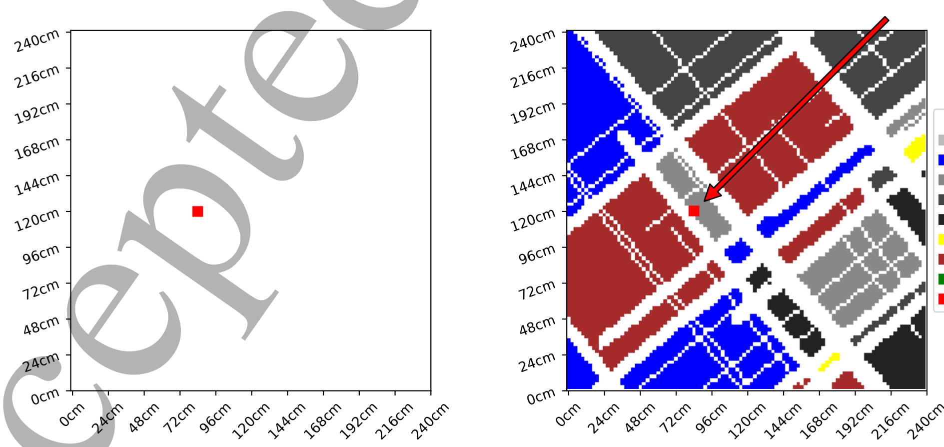

Figure 15. Left: Backprojection with source detected. Right: Cargo configuration with source configuration indicated by arrow. 100,672 particles detected, 99,683 background particles and 989 source particles. Exposure time is 21.92 seconds.

6.3. Performance on Large Scale Dataset

To test the statistical performance of the CNN on a large scale, 1738 unique cargo configurations are generated using an alternate (to avoid possible inverse crime) generative scheme. For each cargo configuration all four linear detector arrays are present, from zero up to four sources are randomly placed and simulated independently, so that by using linearity of (2) we can produce $1738 \times 16 = 27808$ testing samples. Particle detections are simulated for an exposure time measured by the expected background detection levels of 20000, 50000 and 100000 particles. The data were fed into the trained CNN for source presence detection. The results for presence detection are summarized in Table 3 below. The results obtained clearly confirm our expectations (see Section 1).

Expected Particle Count	Sensitivity	Specificity
100000	99.90%	99.71%
50000	99.78%	94.59%
20000	99.81%	36.36%

Table 3. Sensitivity and specificity of the CNN source detection with each source having 1% SNR.

We remind the reader that sensitivity, or true positive rate, shows the success of determining the presence of a source (i.e., few false negatives), while specificity reflects how well the absence of the source is detected (i.e., few false positives). High specificity was hardwired into the BP techniques [5, 6], it was only the sensitivity that was questionable.

The accuracy of the prediction generally increases with particle count (and thus observation time), and sufficient particle counts are required for successful detection. At the low levels (e.g., of 20000 particles and lower) the network seems biased to think that a source is always present. This clearly leads to near 100% sensitivity and an extremely low specificity, which makes the detection practically not feasible, due to high level of false positives.. An explanation could be that the features that are being detected (albeit we do not know what they are) are non-smooth, vs. large smooth background. When the total count is low, the whole dataset becomes non-smooth, which tricks the network.

For comparison, we show below the analogous backprojection results, which are significantly worse. This is not surprising, since the basic assumptions for this technique are not satisfied. For 10^5 particles CNN succeeds extremely well and beats hands down the backprojection technique, which often does not show any statistically significant deviations and thus does not detect presence of the source. Notice that six times higher number of detected particles was required in [5, 33] for backprojection stable detection, even without complex cargo being involved.

Expected Particle Count	Sensitivity	Specificity
100000	71.04%	99.31%
50000	65.55%	99.31%
20000	52.13%	98.91%

Table 4. Sensitivity and specificity of the backprojection source detection with each source having 1% SNR.

6.4. Number of sources and number of detector arrays

Here we address the question of whether one has to completely surround the object with four detectors, or some results can be achieved with three, two, or one flat detector arrays. We thus have simulated each cargo configuration with zero to four independent sources randomly placed. Just as with the training data, we take several combinations of which sources and detector arrays are present. The combinations are summarized in Table 5 below.

Number of Sources	One Detector	Two Adjacent Detectors	Two Opposite Detectors	Three Detectors	Four Detectors	Total
0	6952	6952	3476	6952	1738	26070
1	27808	27808	13904	27808	6952	104280
2	41712	41712	20856	41712	10428	156420
3	27808	27808	13904	27808	6952	104280
4	6952	6952	3476	6952	1738	26070
Total	111232	111232	55616	111232	27808	417120

Table 5. Number of testing samples in each category

Particle detections are simulated for an exposure time measured by the expected background detection levels of 20000, 50000 and 100000 particles. The data were fed into the trained CNN for inference. The results for presence detection are summarized in the bar graphs below, with detailed tables posted in [8].

Additionally, we investigated the effect of scaling the strength of each source so that altogether they had the same strength as a single source, thus effectively diluting the localized signature of the source. In the case of backprojection the localized nature of the source is the key justification for the method of [5]. This would lead one to believe that splitting the source strength will make it more difficult for the CNN to detect any source presence, which is indeed confirmed by the results summarized in Table 6 below.

6.5. Observation time

The above results are presented in terms of the total number of particles detected. The conclusion is natural: the larger - the better. The number of detected particles obviously

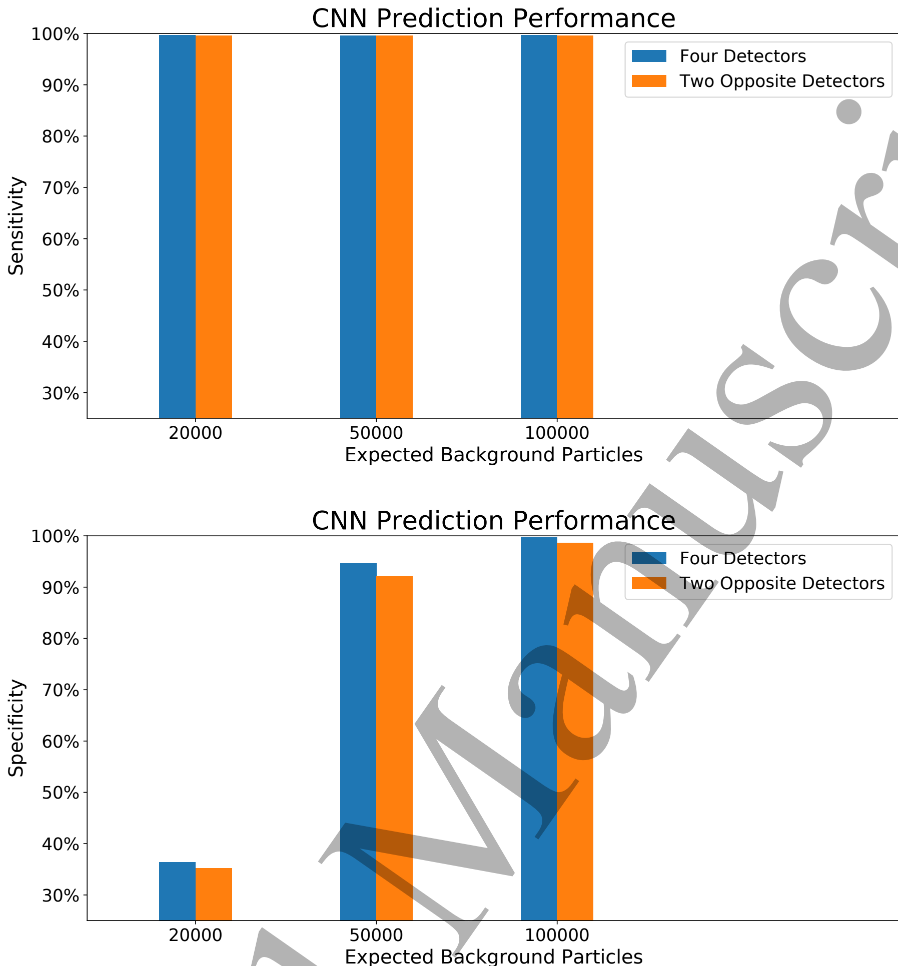

Figure 16. Sensitivity and specificity of the source detection.

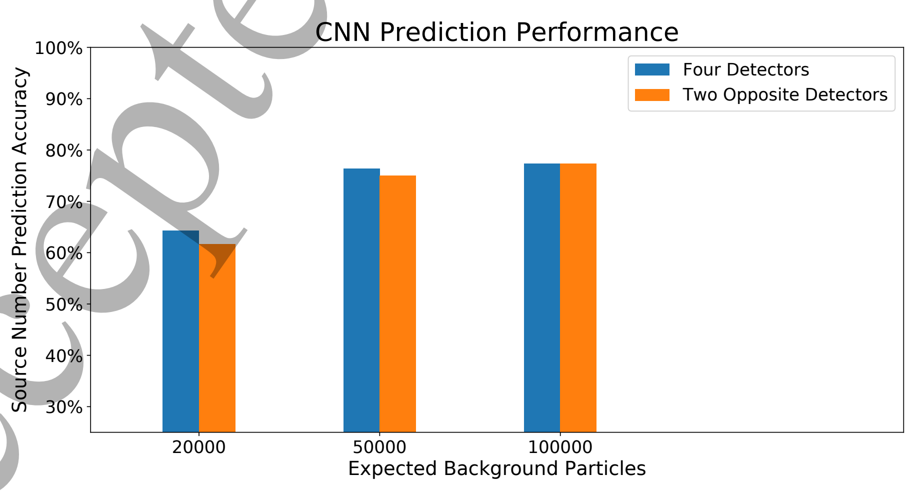

Figure 17. Accuracy of the number of source detection.

Expected Particle Count	Sensitivity one source	Sensitivity two sources	Sensitivity three sources	Sensitivity four sources	Specificity
100000	99.74%	96.57%	89.67%	82.62%	99.71%
50000	99.68%	97.25%	93.51%	89.13%	94.59%
20000	99.99%	99.95%	99.86%	99.77%	36.36%

Table 6. Sensitivity and specificity of the source detection techniques with split source strength.

increases with (essentially proportional to) the time of observation. However, the slope of this increase clearly depends significantly on the type and configuration of the cargo. Thus, the exposure time required to reach a certain level of particle detections is a function of the configuration of the cargo, including source location, material composition, material placement, and background strength. This makes it difficult to predict boundary flux rates, even if the configuration is known, without solving the Boltzmann equation (2).

To make a fair numerical experiment, many heavily iron (and thus very shielding) cargo scenarios have been included. Namely, the set of all samples have been divided into 24 sets of equal size, and the probability of choosing iron as the filling of boxes was increasing linearly from zero in the first group to almost one in the 24th one. Figure 18 contains the histogram of the number of runs vs. time required for detection for thousands of configuration runs for detecting the presence of a source emitting on the order of 1000 particles (assuming four detectors). The vast majority would require time measured in seconds.

Generally speaking, one would epect the large bin on the left-hand side to correspond to configurations with less high-Z materials, and the larger bins on the right-hand side correspond to configurations with more high-Z materials. It can certainly become unrealistic to detect many source particles in some of the latter cases. Nevertheless, as is evidenced by some of the examples presented, as well as statistics presented in Section 6.3, quite a few configurations of high-Z material exist where presence of source(s) source can be detected in a reasonable amount of time. These lower exposure time scenarios would be the most appropriate cases for detecting illicit nuclear materials at border crossings. Some of the longer exposure times (on the order of several minutes to perhaps several days) would be appropriate for detection of illicit nuclear materials in shipping containers on cargo ships, where scanning can be done while the container is in transit.

Additionally, it is important to note that if one restricts oneself to a smaller number of detector arrays (incomplete view), it will take longer to reach the same exposure level and thus would add to the number of undetected cases.

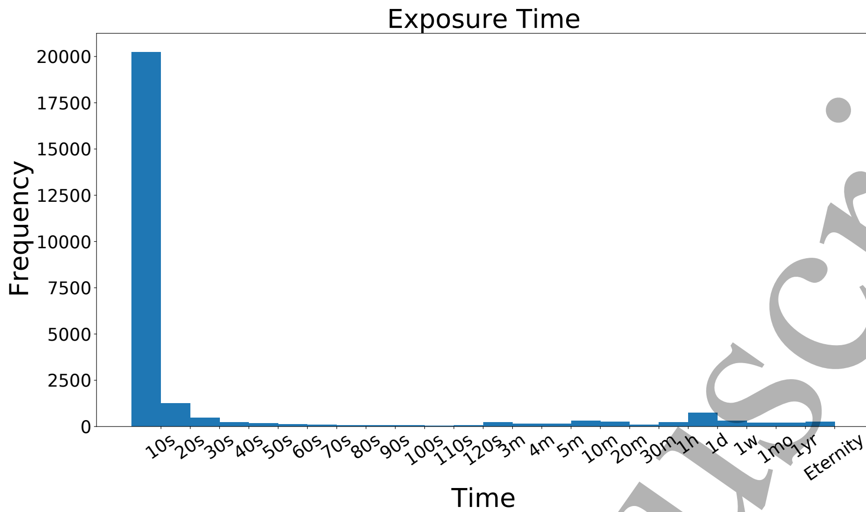

Figure 18. Histogram of the number of runs vs. exposure times required for detection for the testing data set. These times are computed in the case that all four linear detector arrays are present and anywhere between one and four sources are present.

7. Remarks and Conclusions

- Our work shows that the deep learning approach significantly improves over detection by backprojection techniques of [6,30,33] and works for complex attenuating and scattering cargo scenarios, where the latter fails completely. This confirms the opinion expressed in [6] that some information about source presence was there.
- This article concentrates on the cases of presence of complex cargo and much (an order of magnitude) lower number of γ-photon detected than in [5, 33]. This makes the backprojection detection algorithm of these works not only weak, but also groundless.
- The network performs well detecting the number of up to four sources (although naturally somewhat less successfully than detecting mere presence of a source).
- The authors want to make clear that when producing the results of this paper, no processing (e.g., backprojecting) of the raw detector data is done before feeding it to the network. Since the authors do not know what features would be of importance, we have decided to not impose our prejudices on the data (especially taking into account that backprojection is a smoothing operator, and the relevant information is most probably contained in some sharper features).
- The exposure time required for detection is discussed in Section 6.5. The histogram in Figure 18 shows detection in a matter of second for a vast majority of configurations. It is clear that there are some unbeatable shieldings, so one cannot aim for the 100% success rate. In such cases, other detection techniques could be used: from methods of detecting presence of significant amounts of (shielding) high-Z materials, to neutron emission detection, to human intelligence.

- A strong effort has been made to avoid committing an inverse crime. The testing samples have been produced by an algorithm independent of the one used for the training data. The testing cargo geometries were different from the ones not encountered in the training data, so there was no intersection between the two data-sets.
- A variety of symmetry rules, including rotational symmetry and mirror symmetry were applied randomly to some of the configurations and their material content, to check whether presence or absence of the symmetry influence the detectability. The network performance does not seem to react to this.
- The reader should not think that retraining was needed for different tasks and situations, e.g. for heavy iron cargo, or for detecting the number of sources, rather their mere presence. This all was done with a single trained network.
- Four planar Compton detectors forming a square surrounding the object of interrogation were assumed. It seems that this is the most practical design of such detectors. Effects of removal of some of the detectors have also been studied (Section 6.4).
  The rectangular shape causes some problems, though, e.g. in backprojection method they create (easily removable) corner artifacts. More importantly, this design lacks full rotational invariance, which could be beneficial for the NN design. On the other hand, the rectangular case is challenged by appearance of tilted cargo structures in the test samples, while they were absent in the training data. The network, however, clearly has overcome this difficulty.
- Deep learning techniques have been applied for SPECT image reconstruction, but as it has been mentioned in the text the level of SNR we are dealing with in this work makes any attempt to image reconstruction rather than binary detection impossible.
- There are various further improvements that one should attempt (and are being attempted). Some of them are addressed below.
  (i) It would have been great to figure out what specifically were the signs of presence of the source that the network has learned. This would open a door for developing more analytic methods. However, at this moment the authors do not know what these features are.
  (ii) Producing many more training data is a serious stumbling block in 2D, and especially in 3D case.
  (iii) The CNN architecture should be improved, aiming to reach shorter observation time and even lower SNR levels.
  (iv) We are working on moving to the more realistic 3D situation. The significant difference here is, first, the much higher dimensionality of the data (5D) and corresponding much more massive computations that are needed. Second, in 3D, unlike 2D (where a cone consists just of two rays), the Compton data differ significantly from the usual Radon ones. In particular, an issue arises of how to bin the five-dimensional Compton data in such a way, that the use of CNN could be warranted.

(v) The neural network (NN) approach should be tested on real data, which the authors clearly do not have. However, the radiative transport forward computations we used are commonly practiced in nuclear engineering, seem to be very realistic, and involve realistic material parameters. There is a chance that when novel neutron detectors that are being developed are deployed, we could get some real data.

(vi) The approach we describe indicates presence of a source, but not its location (at least in the heavy iron cargo case). One wonders whether location can also be attempted.

(vii) Although the results presented have been obtained by the same once trained NN, during research various designs of the NN and training sets have been experimented with, all showing consistent ability of detection. It would be still important to study further the model uncertainty (e.g., by using the dropout technique [14]). This will be done in a future work.

Meanwhile, although the testing samples often deviated from the structures used in the training set, our results have shown that the NN generalized extremely well. High experimental levels of the sensitivity and specificity, as well as more detailed information presented in Section 6.3 about statistical spread of the results instill confidence in the suitability of the network as a detection tool.

The imperfection of Compton camera detections has been partially addressed by randomizing the source strength and location and finite bin sizes for the detected data. Depending on the quality of the future detectors, the bin sizes might have to be increased and new study conducted.

(viii) When source particles scatter they lose energy. If source particles downscatter to lower energy groups we will lose them in our data since we only use the highest energy group. It would be interesting to try and use these lower energy groups in either a 3D convolution with 1 channel or a 2D convolution with multiple channels to see if we can get better results.

8. Acknowledgements

The first two authors acknowledge the support from the National Science Foundation through the DMS grant #1816430 and Texas A&M Cyclotron Institute. The third author has been partially supported through a grant by the Department of the Defense, Defense Threat Reduction Agency under Award No. HDTRA1-18-1-0020. The content of the article does not necessarily reflect the position or the policy of the federal government, and no official endorsement should be inferred.

The authors are truly indebted to the three referees, whose detailed comments served not only to improve exposition, but even more importantly, attracted our attention to some features we have missed.

References

[1] Abadi M, et. al. 2015 Tensorflow: Large-Scale Machine Learning on Heterogeneous Systems `https://www.tensorflow.org/`

[2] Adams M P, Adams M L, Hawkins W D, Smith T, Rauchwerger L, Amato N M, Bailey T S and Falgout R D 2013 Provably optimal parallel transport sweeps on regular grids *International Conference on Mathematics and Computational Methods Applied to Nucl. Sci. & Eng.* **4** 2535

[3] Adams M P, Adams M L, McGraw C N and Till A T 2015 Provably optimal parallel transport sweeps with non-contiguous partitions *ANS MC2015-Joint International Conference on Mathematics and Computation (M&C), Supercomputing in Nuclear Applications (SNA) and the Monte Carlo (MC) Method* **2** 1218

[4] Adams M P, Adams M L, Hawkins W D, Smith T, Rauchwerger L, Amato N M, Bailey T S, Falgout R D, Kunen A and Brown P 2020 Provably optimal parallel transport sweeps on semi-structured grids *Journal of Computational Physics* **01/2020** 109234

[5] Allmaras M, Darrow D, Hristova Y, Kanschat G and Kuchment P 2010 Detecting small low emission radiating sources *Inverse Problems & Imaging*, **7(1)** 47

[6] Allmaras M, Ciabatti A, Hristova Y, Kuchment P, Olson A and Ragusa J 2016 Passive Detection of Small Low-Emission Sources: Two-Dimensional Numerical Case Studies *Nuclear Sci. and Eng.* **184(1)** 125

[7] Vegard Antun, Francesco Renna, Clarice Poon, Ben Adcock, Anders C. Hansen, On instabilities of deep learning in image reconstruction and the potential costs of AI, Proceedings of the National Academy of Sciences May 2020, 201907377; DOI: 10.1073/pnas.1907377117

[8] Baines W., Kuchment P, Ragusa J, Preprint arXiv:2003.04089.

[9] Bell G I and Glasstone S 1970 *Nuclear reactor theory.* (Malabar, FL: Krieger Publishing Company)

[10] Carlson B G 1955 *Solutions of the Transport Equation by SN Approximations* (Los Alamos, NM: Los Alamos Scientific Laboratory)

[11] Chen G, Esch G, Wonka P, Müller P and Zhang E 2008 Interactive Procedural Street Modeling *ACM Trans. Graph.* **27** 1

[12] Chollet F 2015 Keras *GitHub* `https://github.com/fchollet/keras`

[13] Duderstadt J J and Martin W R 1979 *Transport theory* (New York, NY: John Wiley & Sons)

[14] Y. Gal, Z. Ghahramani, Dropout as a Bayesian approcimation: representing model uncertainty in deep learning, Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 2016. JMLR: W&CP volume **48**.

[15] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press)

[16] Hawkins W D, Bailey T S, Adams M L, Brown P N, Kunen A J, Adams M P, Smith T, Amato N M and Rauchwerger L 2014 Validation of Full-Domain Massively Parallel Transport Sweep Algorithms *Trans. Amer. Nucl. Soc.* **111** 699

[17] Jones E, et. al. 2001 SciPy: Open Source Scientific Tools for Python `http://www.scipy.org/`

[18] Kingma D P and Ba J 2015 Adam: A Method for Stochastic Optimization *International Conference on Learning Representations*

[19] Lamarsh J 1966 *Introduction to nuclear reactor theory.* (Boston, MA: Addison-Wesley series in nuclear engineering. Addison-Wesley Pub. Co.)

[20] Lesaint P and Raviart P A 1974 On a finite element method for solving the neutron transport equation. *Mathematical Aspects of Finite Elements in Partial Differential Equations*, **33** 89

[21] Lewis E E and Miller W F 1984 *Computational methods of neutron transport* (United States: John Wiley and Sons, Inc.)

[22] Liu W, Wang Z, Liu X, Zeng N, Liu Y and Alsaadi F 2017 A survey of deep neural network architectures and their applications, *Neurocomputing* **234** 11

[23] MacFarlane R and Muir D 1994 *The NJOY nuclear data processing system version 91* (Los Alamos, NM: Los Alamos Scientific Laboratory)

583 [24] Natterer F 2001 *Mathematics of computerized tomography* (Philadelphia, PA: SIAM: Society for Industrial and Applied Mathematics)

585 [25] Perlin K 1985 An Image Synthesizer *Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques* **10** 287

587 [26] Reed W H and Hill T R 1973 Triangular mesh methods for the neutron transport equation. *National topical meeting on mathematical models and computational techniques for analysis of nuclear systems* Los Alamos Report LA-UR-73-479

590 [27] Sánchez R and Ragusa J 2011 On the construction of galerkin angular quadratures *Nuclear Science and Engineering* **169(2)** 133

592 [28] Santi P A 2013 Passive nondestructive assay of nuclear materials (Los Alamos NM: Los Alamos Scientific Laboratory) Los Alamos Report LA-UR-73-479

594 [29] Emil Y. Sidky, Iris Lorente, Jovan G. Brankov, Xiaochuan Pan, Do CNNs solve the CT inverse problem?, 2020, arXiv preprint arXiv:2005.10755

596 [30] Terzioglu F, Kuchment P and Kunyansky L 2018 Compton camera imaging and the cone transform : A brief overview *Inverse Problems* **34(5)** 054002.

598 [31] Wareing T, McGhee J, Morel J and Pautz S 2001 Discontinuous finite element Sn methods on three-dimensional unstructured grids *Nuclear Science and Engineering*, **138(3)** 256

600 [32] Wu K, Otoo E and Shoshani A 2005 Optimizing connected component labeling algorithms *Medical Imaging 2005: Image Processing* **5747** 1965

602 [33] Xun X, Mallick B, Carroll R and Kuchment P 2011 Bayesian approach to detection of small low emission sources *Inverse Problems* **27(11)** 115009

604 ## 9. Appendix: Algorithm for Procedural Generation of Cargo Configurations

Algorithm 1: Procedural Cargo Configuration

```
Generate Perlin noise in cargo;
Initialize n_x and n_y to desired number of vertical and horizontal boundaries
  (numbers can be chosen randomly);
Sum Perlin noise over rows and columns to produce noise function on edge of cargo;
Randomly select n_x distinct x−coordinates for vertical boundaries and n_y distinct
  y−coordinates for horizontal boundaries according to edge noise functions. Store
  in x and y respectively.;
n_iter = 0 ;
while n_x > 0 or n_y > 0 do
    if n_iter is even and n_x > 0 then
        Determine all existing boundary points along the line (x[n_iter/2], y).
          Randomly select a starting point y_s and ending point y_e from among the
          existing boundary points according to previously generated Perlin noise.
          Set all points between (x[n_iter/2], y_s) and (x[n_iter/2], y_e) to boundary
          points.;
        n_x = n_x − 1;
        n_iter = n_iter + 1;
    else if n_iter is odd and n_y > 0 then
        Determine all existing boundary points along the line (x, (y[(n_iter − 1)/2]).
          Randomly select a starting point x_s and ending point x_e from among the
          existing boundary points according to previously generated Perlin noise.
          Set all points between (x_s, (y[(n_iter − 1)/2]) and (x_e, (y[(n_iter − 1)/2]) to
          boundary points.;
        n_y = n_y − 1;
        n_iter = n_iter + 1;
end
Identify connected components (Scipy.Measure.Label);
if Rotational Symmetry Desired then
    Copy one quadrant of the configuration over all others with appropriate
      rotation;
if Mirror Symmetry Desired then
    Copy one side of the configuration over the other with mirroring ;
... Randomly assign material identification to each connected component ;
Save configuration to file;
Result: Single cargo configuration
```