Journal of Transportation Security (2020) 13:71–89
https://doi.org/10.1007/s12198-020-00211-5

CH-Net: Deep adversarial autoencoders for semantic segmentation in X-ray images of cabin baggage screening at airports

Mohamed Chouai[1] · **Mostefa Merah**[1] · **Malika Mimi**[1]

Received: 2 January 2020 / Accepted: 25 June 2020 / Published online: 6 July 2020

Abstract
Billions of suitcases and other belongings are checked every year in the X-ray systems in airports around the world. This process is of great importance because it involves the detection of possible dangerous objects such as weapons and explosives. However, the work done by airport surveillance personnel is not free from errors, usually due to tiredness or distractions. This is a security problem that can always be reduced with the help of automatic intelligent tools. This paper attempts to make a contribution to the field of object recognition in X-ray testing for luggage control by proposing a deep learning system that combines a deep convolutional network with an adversarial autoencoder acting as a powerful feature extractor mechanism. The system is developed to separate transmission X-ray images into potentially overlapping regions, separating the X-ray image into organic and inorganic images, taking into consideration the overlapping between the same and different types of materials. To show the superiority of our proposed system, a comparative analysis was carried out including the state-of-the-art deep learning semantic segmentation systems. The proposed method demonstrated highly promising results, achieving the best performance in global accuracy, mean boundary F1 and mean IoU, with a percentage of 80.17%, 76.28% and 76.84%, respectively.

Keywords Transport security · Baggage control · Semantic segmentation · Deep learning · Autoencoder

✉ Mohamed Chouai
mohamed.chouai@univ-mosta.dz

[1] Signals and systems laboratory, Department of Electrical Engineering, Mostaganem University, Site 1 Route Belahcel, 27000 Mostaganem, Algeria

Introduction

Since September 11, 2001, air safety by X-ray scanners has become more than important. The control process is complex, since threatening objects are very difficult to detect when they are placed in closed bags, superimposed by other objects and / or oriented in different positions, thus making their detection difficult. In baggage control, where the complexity of the inspection is very high, human operators are still used. However, they only have a few seconds to decide whether a bag contains a prohibited item or not, thus, the detection performance hardly exceeds 90% (Michel et al. 2010). Although, the human operator remains essential in this type of technology, it presents a certain number of risk factors such as distraction or inattention errors, which constitutes a major risk, hence a decision aid by computer is required.

The building materials used in a typical travel bag can be derived from plastic, fabric, leather, synthetic or natural fabrics, metal frames or a mixture of these. However, for inspection personnel, the contents of the bag are of primary importance because, associated with the material and construction of the bags and luggage, creates a background scene in which potential targets may be present. In this regard, it is virtually impossible to make assumptions about the content of cabin baggage. Apart from the various prohibited items such as weapons (firearms), ammunition, various foodstuffs, sensitive material (for example pornography) that different border agencies apply, we cannot make any assumptions about the contents of the luggage cabin. This has been exploited by terrorist groups, so the explosives can often appear, upon X-ray inspection, to look like everyday objects. For example, dynamite can have the appearance of a marzipan bar.

In baggage screening, the X-rays pass through a suitcase and are then attenuated (weakened) by the different densities of objects inside the bag that they encounter. An enlarged "shadow" of the bag and its contents is then displayed on the computer screen of the X-ray machine. X-ray photons penetrate most materials. As a result, all objects along an X-ray path attenuate the electromagnetic radiation and contribute to the measured final intensity. This is the transparency property which allows X-ray imaging to see through objects. More importantly, unlike reflection images in which each pixel corresponds to a single object, the pixels of the transmission images represent the attenuation of several objects. For regular images, artificial vision approaches decompose an image into disjunct regions (segments) that must roughly correspond to the objects. For radiological images, however, pixels should not be assigned to a single region, but to all overlapping objects that contribute to their attenuation.

Baggage screening is a difficult process that requires a significant commitment of personnel and resources to detect threats that can be hidden among the millions of bags that are carried through airports daily. All airport baggage security technologies have one thing in common: the final decision as to whether a suitcase will be carried in an aircraft remains in the mind of a human operator. Although advances in security technologies are crucial, all screening equipment must be used with highly qualified and motivated personnel.

Several challenges are involved in this area, including homemade explosives and their precursors. The variation in the density of homemade explosives (HME) often depends significantly on the manufacturing process. The makers of the

illegal bombs seem to have a good understanding of how X-ray scanning equipment works, and where possible they would likely seek to adapt their technology so that the device becomes undetectable, particularly with respect to familiar X-ray transmission imaging systems.

Faced with increased risks, airports and governments have increased their investments in aviation security (Gillen and Morrison 2015). Over the past five years, explosive detection systems for cabin baggage (EDSCB) have been available (Sterchi and Schwaninger 2015). There are few countries such as the United States that use these systems (Peter 2015).

In recent years, deep learning has been used successfully in image and video recognition (Najafabadi et al. 2015; Chouai et al. 2019). The goal affiliated with our team is to develop a system using machine learning and advanced learning to effectively detect threatened objects (knives, firearms, explosives).

Applying artificial vision methods to X-ray baggage images has many advantages over manual inspection, including improved performance and high inspection throughput due to the automation. Few works have been published on the automated processing of X-ray images of baggage, probably due to the difficulty of obtaining sufficiently large labeled data sets and also due to the confidentiality of its data.

An accurate and reliable automatic detection of objects at risk is more than necessary. However, this type of detection suffers from two major problems, in which we will try to solve them in this article: i) the detection of objects at risk, masked or overlapped by other objects; and ii) the rate of false alarms, which by force can hinder the use of this type of system.

In this paper, we develop a method to separate transmission images into potentially overlapping regions. In particular, the X-ray images are separated into organic and inorganic images, taking into consideration the overlapping between the same and different types of materials. The term separation distinguishes our output from traditional segmentation where each pixel belongs to a single region. To address this problem, we label each pixel of the image as belonging to one of six different classes. This assignment is made according to the context of the different objects of the image using the semantic segmentation which has never been used in this domain. This paper attempts to make a contribution to the field of object recognition in X-ray testing for luggage control by proposing a Semantic Segmentation by Adversarial AutoeEcoders, one of the most promising generative learning machines (Makhzani et al. 2015) (throughout the article, the proposed system will be named as CH-Net). The performance provided by the proposal system will be compared with those obtained by different deep learning semantic segmentation strategies that have been recently proposed. The object detection process will follow the segmentation task by applying the labeling technique of the related components (Sezgin and Sankur 2004; Luigi and Andrea 1999) and some fusion techniques imposed by us (will be detailed in Section "Data preparation").

This paper is structured as follow. Section "Image segmentation for X-ray baggage images" presents a review of the segmentation process in the domain of baggage inspection. A brief description of the semantic segmentation and its domain application is introduced in Section "Semantic segmentation". In Section "Proposed system (CH-Net)" the phases involved in our proposed system, the background

information with the implementation and the data preparation details are explained. A comparative analysis showing the description of the used algorithms is presented in Section "Comparative analysis". Section "Results and discussion" shows the experimental results with the discussion and interpretation. Finally, Section "Conclusion" closes the article with conclusions and future work recommendations.

Image segmentation for X-ray baggage images

Image segmentation is at the heart of many imaging problems since it is often the first and the most important step in image processing. Most of the automatic object detection systems are based on a segmentation procedure consisting of an image processing to gather pixels together according to pre-defined criteria. The segmentation of scanned X-ray images for baggage control has taken place over the past two decades. A brief summary of the most relevant existing methods for segmentation of X-ray images of baggage is presented next.

Mery et al. (2017) studied the object recognition applying different modern of computer vision techniques. Their images were segmented using an adaptive k-means clustering method. Wiley et al. (2012) adapted their Stratovan Tumbler (Wiley 2012) segmentation method to automatically segment objects from Computed Tomography (CT) scanned images. Their method is based on the use of a kernel that virtually moves in three dimensions, the kernel starts inside the object and moves iteratively in all directions until it reaches the boundary of the objects. A specific method for weapon detection is proposed in Abidi et al. (2006). In this work, Abidi *et al.* present a segmentation procedure for single energy X-ray images through pseudo colouring and using a multilevel thresholding segmentation scheme.

DeDonato et al. (2014) presented an automated checking baggage inspection system with robots. They used the Point Cloud Library (PCL) to implement plane segmentation (Rusu and Cousins 2011) allowing them to remove extra surfaces from the image. Hassanpour (2016) developed a material detection baggage system where the image is segmented into sub-regions based on the intensity level of pixels, followed by an overlapping detection algorithm in order to estimate the regions belonging to each intensity layer. Furthermore, Mery (2014) worked on computer vision technology for X-ray testing. They used histograms, edge detection, morphological operations and filtering for the segmentation phase. In addition, Megherbi et al. (2010) proposed a system for detecting potential threats in CT baggage screening. They used medical approaches segmentation, which consists of region-based methods (which perform segmentation by finding coherent regions according to certain criteria) and boundaries-based methods (which find the boundaries of objects of interest). Also, in 2013 Megherbi et al. (2013), the same authors studied the application of existing CT scanning segmentation techniques to the automated control of baggage and packages; in particular, they explored and focused on confidence connected region growing, fuzzy connectedness, watershed, and fast matching.

Liang et al. (2003) proposed an automatic X-ray image segmentation method for low-intensity threat detection. They used multi-thresholding and data clustering

techniques for the segmentation process. Martin et al. (2015) proposed a learning-based framework for joint segmentation and identification of objects in Dual-Energy X-ray Computed Tomography (DECT) images. They focused on segmenting and identifying a small set of objects of interest and considering everything else as background. Similarly, Babaheidarian and Castañón (2017) focused on the problem of identifying materials in DECT images. They proposed a new algorithm for joint segmentation and classification of material regions. In addition, Muthukkumarasamy et al. (2004) aimed to develop an intelligent banned object detection system to enhance aviation security by converting each image to grayscale after segmenting them using a multiple-thresholding algorithm. Heitz and Chechik (2010) introduced a method for separating objects in a set of X-ray images using the additivity attenuation property at the pixel level (the sum of the attenuations of all objects traversed by the X-rays). This method exploits multiple projection views of the same scene from slightly different angles to produce an accurate estimate of the attenuation properties of the objects in the scene.

Semantic segmentation

Semantic segmentation is a basic building block for image comprehension. By classifying all the pixels of an image in a dense way, it is then possible to construct interesting abstract representations in the objects and their forms. Its applications include image understanding, autonomous driving, object recognition, machine translation, and machine vision. Semantic segmentation is a very active area of research because of its importance in real-world applications (Romera et al. 2017; Zhao et al. 2018; Mehta et al. 2018; Siam et al. 2018). The goal is to classify each pixel in the image into a specific predefined category, *i.e.*, it is intended to label each pixel in the image as belonging to one class. This assignment is made according to the context of the different objects of the image.

Proposed system (CH-Net)

In this work, a new semantic segmentation model based on Adversarial AutoEncoder (AAE) is presented. This method is thought to be highly efficient in terms of memory, operation, computational time and competitive with the main state-of-the-art segmentation procedures. In this way, the proposed architecture is designed to train an end-to-end pixel-label classifier by means of a stochastic gradient descent and the cross-entropy loss function. A scheme of the CH-Net is presented in Fig. 1.

AAE is similar to an autoencoder, but the encoder is adversely trained to force it to produce a specific latent space distribution. AAE has a direct relation with Generative Adversarial Networks (GANs) which arranged with three main blocks: the encoder, the decoder and the discriminator (Makhzani et al. 2015). The encoder takes in an input **x** and produces an output **h** (latent space) by compressing the image such that it will occupy a lower dimensional space. The decoder takes the **h** and tries to recon-

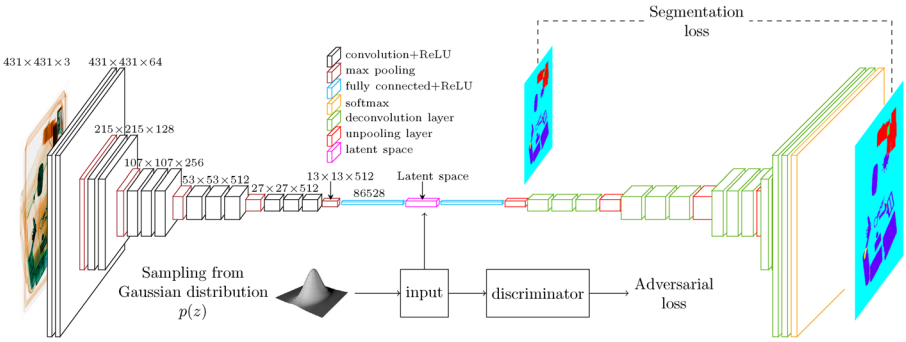

Fig. 1 Architecture of the proposed system (CH-Net). The encoder of the AAE is identical to the convolutional layers in VGG16 (Simonyan and Zisserman 2014). It consists of convolutional layers of 3×3 filters. The decoding network replicates the encoding process in inverse order. The decoder networks ends with a pixel classifier with a softmax output layer

struct the input at its output by minimizing the reconstruction loss between the original image and the output image of the decoder . The discriminator to discriminate between a samples generated using a prior distribution and sample from hidden code.

Training an AAE has 2 phases: reconstruction phase and the regularization phase. In the first phase, both the encoder and the decoder are trained to minimize the reconstruction loss; in the second one, the discriminator is trained to distinguish between the encoder outputs and random inputs, and the generator is trained to force the encoder to output latent code with a desired distribution (Makhzani et al. 2015).

We have chosen the AAE to learn the features latent space because it can produce feature vectors with a desired distribution. In this work, the AAE is trained in order to produce a symmetric Gaussian distribution of the feature space. In this way, all the features in the latent space are uncorrelated, *i.e.*, each feature captures a unique and specific characteristic of the input image. In other words, the AAE acts as a powerful feature extractor.

Proposed architecture

Our segmentation approach follows a symmetric encoder-decoder CNN architecture using an AAE network. The encoder of the AAE is identical to the convolutional layers in VGG16 (Simonyan and Zisserman 2014). It consists of convolutional layers of 3×3 filters, repeated several times changing the number of kernels. For the purpose of robust feature extraction, batch normalization (Ba et al. 2016) and ReLu (Agarap 2018) are used after the convolution operation. The decoding network replicates the encoding process in inverse order. The decoder networks ends with a pixel classifier with a softmax output layer. This architecture is shown in Fig. 1.

The network is trained on an NVIDIA GeForce GTX 1050ti GPU. We stop the training when the segmentation loss no longer decrease.

Data preparation

The images used in this study came from the Hi-Tech Detection Systems Society (HDTS) private database with 2000 dual-energy X-ray luggage scan images with a mean size of 800px by 800px. To solve the problem of overlapping objects, a manual segmentation was prepared labeling each pixel as belonging to one of six classes: organic objects, inorganic objects, overlapping between organic/organic, overlapping between inorganic/inorganic, overlapping between organic/inorganic and the background.

The object detection process is implemented after the segmentation task, performing this three fusion rules :

1. An overlapping between two objects of the same material is considered as belonging to both of them.
2. An overlapping between two different materials is considered as belonging to both of them.
3. The background will be ignored.

After this operation we can get two masks, one for organic objects and the other for inorganic objects, by applying the labeling technique of the related components (Sezgin and Sankur 2004; Luigi and Andrea 1999) (performs two-dimensional connected component labeling by analyzing the adjacency relationships of runs of pixels, on the other hand, analyzes the adjacency relationships of individual pixels) we can detect all the presented objects in the luggage.

Comparative analysis

To show the performance of the CH-Net, a comparative analysis is carried out to five state-of-art models deep learning semantic segmentation algorithms, which are described in the following sections.

Fully Convolutional Networks (FCN)

The FCN network is an extension of the classical CNN. The main idea is to make the classical CNN take as input arbitrary-sized images (Long et al. 2015). FCN is built only from locally connected layers, such as convolution, pooling and upsampling. This reduces the number of parameters and calculation time. The main contributions of the FCN are the popularization of the use of end-to-end convolutional networks for semantic segmentation and the introduction of skip connections to improve the coarseness of upsampling. There are variants of the FCN architecture, which mainly differ in the spatial precision of their output. For our application, we will use FCN-32s, FCN-16s and FCN-8s variants. Figure 2 shows the model architecture.

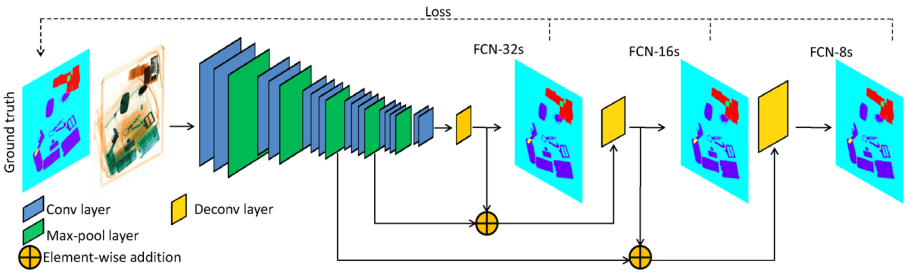

Fig. 2 Architecture of the FCN network

SegNet

SegNet is a deep convolutional encoder-decoder architecture. The encoder applies convolution, batch normalization, and a max-pooling nonlinearity operation, while storing the index of the value extracted from each window. Decoder is similar to encoder, with the difference that it has no nonlinearity and that oversamples its input using the previously stored indexes at the encoding stage. The model architecture is shown in Fig. 3. After the decoder phase, the output is routed to a final pixelwise classification layer (Badrinarayanan et al. 2017).

U-Net

U-Net is a deep CNN based on an encoder-decoder architecture designed especially for biomedical image segmentation. The architecture shown in Fig. 4 contains two paths. The first path is the encoder which is a traditional stack of convolutional and max-pooling layers. The second path is the decoder (symmetric expanding path) which is used to enable precise localization using transposed convolutions (Ronneberger et al. 2015).

DeconvNet

DeconvNet proposed in Noh et al. (2015) is a deep CNN based on an encoder-decoder architecture that combines bottom-up region proposals with multi-layer

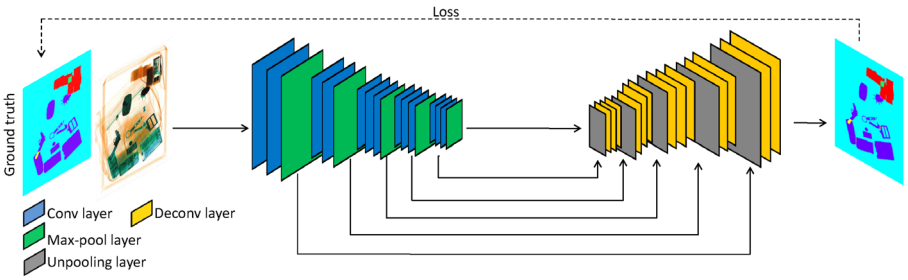

Fig. 3 Architecture of the SegNet network

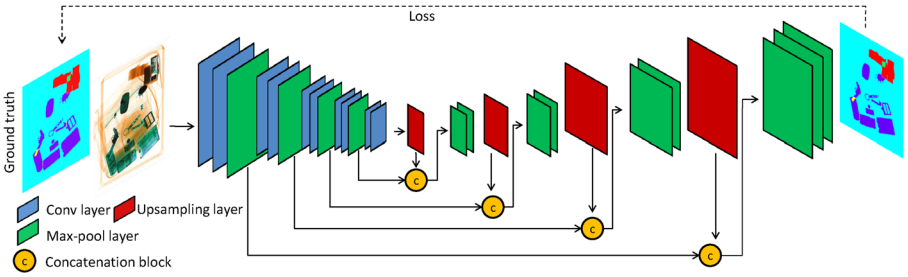

Fig. 4 Architecture of the U-Net network

deconvolution network. The first part is a convolution network which is like that of the FCN, with convolution and pooling layers. The second part is the deconvolution network which is a novel part. The deconvolution network is composed of layers of deconvolution and unpooling layer, which identify class labels in pixels and predict segmentation masks. The overall architecture is shown in Fig. 5.

RedNet

RedNet is an encoder-decoder residual network combines the short skip-connection in the residual unit and the long skip-connection between encoder and decoder for accurate semantic inference (Jiang et al. 2018). The layers are composed by convolution layer, deconvolution layer, and addition. In order not to lose location information and spend extra memory, the authors rid the max-pooling. The architecture of this network is shown in Fig. 6.

Results and discussion

The HDTS dataset introduced in Section "Data preparation" is partitioned into a training set containing 80% of the data (1600 images) and a testing set containing 20% of the dataset (400 images). The comparative results are calculated from the final pixel-level classification of each algorithm. A pixel label is considered correctly

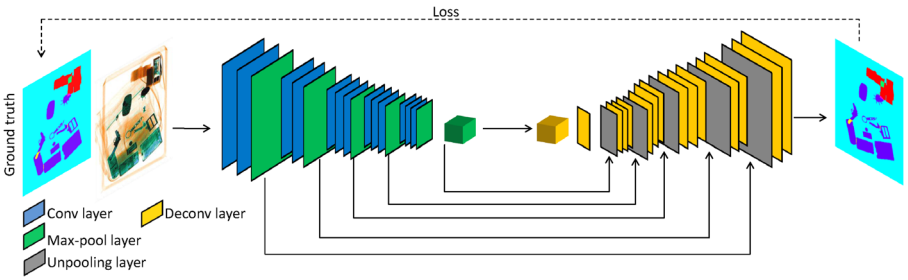

Fig. 5 Architecture of the DeconvNet network

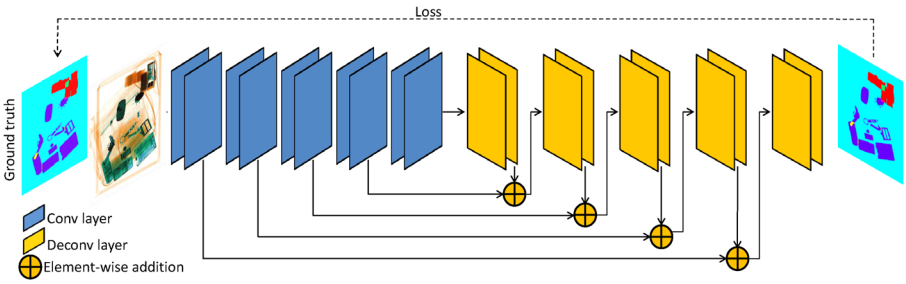

Fig. 6 Architecture of the RedNet network

detected if it is identical to the ground truth label. The performance is measured in terms of accuracy, boundary F1 and pixel Intersection over-Union (IoU) across the six classes.

For the training of the CH-Net, we employed the VGG-16 network which has been pre-trained on ImageNet. We fine-tuned the VGG-16 network to our dataset by minimizing the cross-entropy loss function using a stochastic gradient descent algorithm. We use a mini-batch of six images and an initial learning rate equal to 0.001 which is multiplied by 0.1 at every 800 iterations (a learning rate of 0.01 for the final classifier layer). The momentum and weight decay parameters are established to 0.9 and 0.005, respectively.

For training and testing steps, we resize all the input images and the ground truth semantic maps into a $431 \times 431 \times 3$ resolution. For the testing phase, the image will be returned to its original size after the decoder operation.

The proposed method demonstrated highly promising results. Table 1 lists the performance comparison, in terms of accuracy, boundary F1 and IoU scores, between the proposed system and the tested methods. The average result over the six classes results are shown in Table 2 (high-performance results are shown in bold). These results demonstrate that the CH-Net provides better results than the other tested algorithms showing the effectiveness of the system in the semantic segmentation. In particular, the method obtains 80.17%, 76.28% and 76.84% corresponding to the global accuracy, mean boundary F1 and mean IoU. On the other hand, Table 1 shows how the SegNet provides good results especially in the segmentation of the inorganic objects, and in the segmentation of those objects which overlap with inorganic objects, as shown in bold in the table.

Overall, the semantic segmentation results reveal the difficulty of identifying the organic/organic overlapping class. Among convolutional networks, FCN-32s has the inferior capability for distinguishing this class. Results also confirm the importance of skip connections for semantic segmentation, as leveraging skip connections in FCN-16s and FCN-8s significantly contributed to the improvement of the segmentation results compared to the FCN-32s by retrieving more spatial details. Much detail, however, was obtained from FCN-8s with an improved accuracy of 13.89% (for organic/organic class), due to employing two skip connections relative to FCN-16s. Furthermore, all methods are successful in discriminating the background class.

Table 1 Segmentation result showing the comparative performance between the proposed system and the other tested methods

	Organic	Inorganic	Overlapping organic / organic	Overlapping inorganic / inorganic	Overlapping organic / inorganic	Background
	CH-Net					
Accuracy	**86.83%**	85.32%	**71.76%**	69.80%	**79.43%**	**87.89%**
The boundary F1	**79.42%**	82.76%	**64.08%**	69.62%	**72.11%**	**89.67%**
IoU	**83.65%**	85.92%	**67.21%**	**64.75%**	**73.63%**	**85.86%**
	FCN-32s					
Accuracy	53.15%	57.23%	31.79%	43.71%	47.46%	58.92%
The boundary F1	48.74%	57.67%	36.11%	46.53%	47.14%	56.70%
IoU	59.97%	62.83%	42.24%	37.66%	48.66%	60.89%
	FCN-16s					
Accuracy	54.52%	59.89%	34.45%	46.37%	50.12%	61.58%
The boundary F1	50.11%	60.33%	38.77%	49.19%	49.80%	59.36%
IoU	61.34%	65.49%	44.90%	40.32%	51.32%	63.55%
	FCN-8s					
Accuracy	65.72%	65.06%	39.62%	51.54%	55.29%	66.75%
The boundary F1	61.31%	65.50%	43.94%	54.36%	54.97%	64.53%
IoU	72.54%	70.66%	50.07%	45.49%	56.49%	68.72%
	SegNet					
Accuracy	79.95%	**88.20%**	59.88%	**78.68%**	70.55%	82.01%
The boundary F1	73.54%	**90.64%**	61.20%	**71.50%**	70.23%	79.79%
IoU	81.77%	**87.80%**	65.33%	62.63%	71.75%	83.98%

Table 1 (continued)

	Organic	Inorganic	Overlapping organic / organic	Overlapping inorganic / inorganic	Overlapping organic / inorganic	Background
	U-Net					
Accuracy	59.69%	71.09%	45.65%	57.57%	61.32%	72.78%
The boundary F1	55.28%	71.53%	49.97%	60.39%	61.00%	70.56%
IoU	66.51%	76.69%	56.10%	51.52%	62.52%	74.75%
	Deconvet					
Accuracy	51.86%	58.52%	33.08%	45.00%	48.75%	60.21%
The boundary F1	47.45%	58.96%	37.40%	47.82%	48.43%	57.99%
IoU	58.68%	64.12%	43.53%	38.95%	49.95%	62.18%
	RedNet					
Accuracy	55.87%	61.24%	35.80%	47.72%	51.47%	62.93%
The boundary F1	51.46%	61.68%	40.12%	50.54%	51.15%	60.71%
IoU	62.69%	66.84%	46.25%	41.67%	52.67%	64.90%

Table 2 Average result of accuracy, boundary F1 and IoU between the proposed system and the other tested methods

	Accuracy	The boundary F1	IoU
CH-Net	**80.17%**	**76.28%**	**76.84%**
FCN-32s	48.71%	48.81%	52.04%
FCN-16s	51.16%	51.26%	54.49%
FCN-8s	61.35%	61.46%	64.69%
SegNet	76.54%	74.48%	75.54%
U-Net	57.33%	57.44%	60.67%
DeconvNet	49.57%	49.67%	52.90%
RedNet	52.50%	52.61%	55.84%

This is partially attributed to the availability of a larger number of training samples for background class as compared to the other ones. Theoretically, the accuracies of all classes should be improved upon the inclusion of a greater number of training samples.

The superiority of the CH-Net network compared to state-of-the-art semantic segmentation methods is attributed to the adversely trained manner of the encoder which force to produce a specific latent space distribution, in our case, symmetric Gaussian distribution that produces uncorrelated features. This enriched the contextual information, which is very important for distinguishing different classes. For example, the class of overlapping between organic/organic is the most difficult to discriminate due to high intra-class and low inter-class variance; however, the proposed system obtained an accuracy of 71.76%, as its pixel-based labeling is clear, accurate, and comparable to the ground truth image.

While some image segmentation results are shown in Fig. 7, comparative results between the segmentation obtained by the considered methods and the desired ground truth are shown in Fig. 8. In this figure, the green and magenta regions highlight areas where the segmentation results differ from the expected ground truth (the green color indicates the false positive segmentation and the magenta color indicates the true negative segmentation). Both figures show satisfactory results compared to the other ones. Nevertheless, the overall outcome of the FCN-8s results is the most affected by noisy scatter points, especially around the edges. This can be observed by the granular segmentation at the boundary between the organic objects and the background, this is due to the fact that the network propagates the image through several alternated convolutional and pooling layers, the resolution of the output feature maps is downsampled. Therefore, the direct predictions of FCN are typically in low resolution, resulting in relatively fuzzy object boundaries.

To show the performance of the proposed system for distinguishing between overlapping objects, we use segmented images to detect objects in the luggage. For that, we propose a detection method based on a labeling technique of the related components (Sezgin and Sankur 2004; Luigi and Andrea 1999), after fusing the six masks resulting from the segmentation phase into two, one for the organic object and the

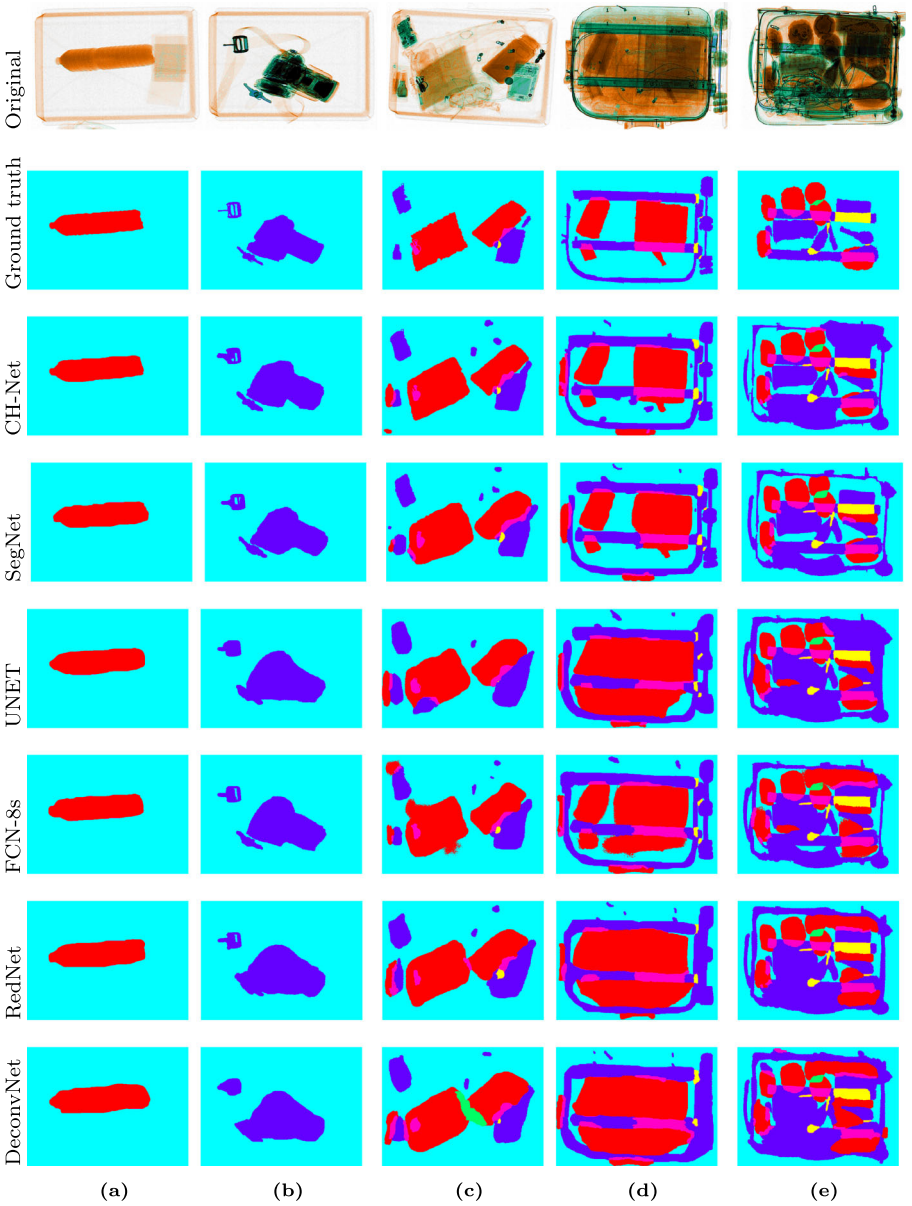

Fig. 7 Some experimental segmentation results over test images. The columns represent the images tested with low to high complexity and the rows represent the different considered methods. It can be observed how the proposed method (third row) produces the most similar segmentation to the desired solution (second row)

other for the inorganic objects according to the rules mentioned in Section "Data preparation". This technique is applied to obtain disjoint sets that can be easily isolated. The model returns the labels of the detected objects, as well as their the

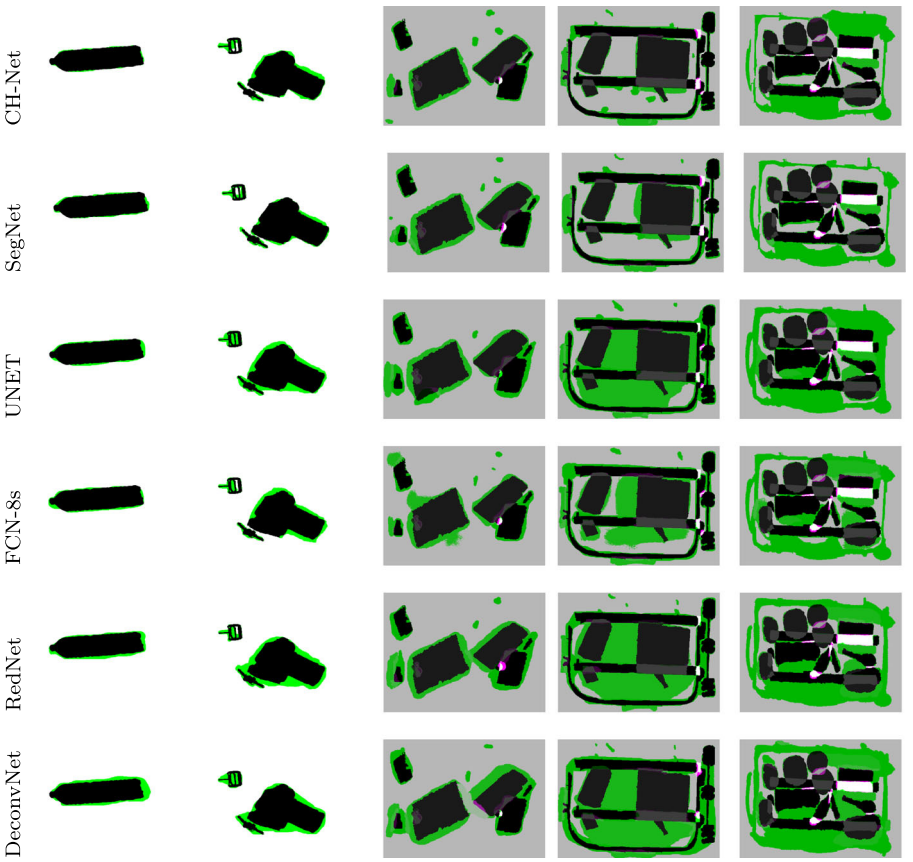

Fig. 8 Comparative results between the segmentation obtained by the considered methods and the ground truth. The green and magenta regions highlight areas where the segmentation results differ from the expected ground truth

coordinates. Using these values, the application generates a new image filled with rectangles surrounding the detected objects, in which the red and the blue boxes represent organic and inorganic objects, respectively. Separate images can be generated for the categories of the organic and inorganic objects, allowing the human operator to discriminate between the selected objects to facilitate the diagnosis. Moreover, the separated objects will be the input of the threat object detection. We can notice that the obtained results are very satisfactory with a small number of false detections as it is shown in Fig. 9, which is very encouraging for future works as, for example, objects classification.

Our study is based on overlapping between two objects, regardless of whether the overlapping is between the same material or different materials. Therefore, there is a limitation in the case of an overlap between three (or more) objects.

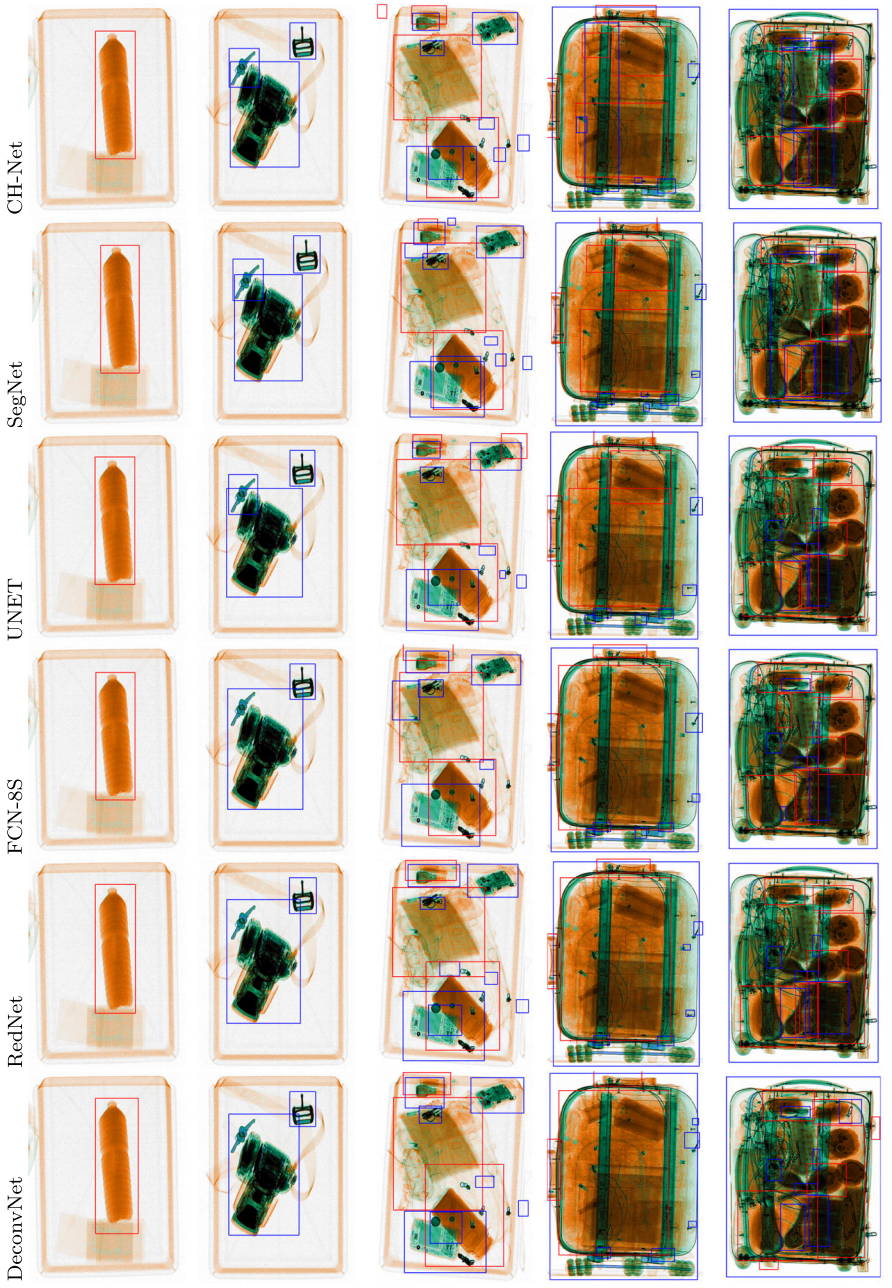

Fig. 9 Object detection example. The red and the blue boxes represent respectively the organic and the inorganic objects. We can notice that the obtained results of the proposed system (first row) are very satisfactory with a small number of false detections. The superiority of the CH-Net is attributed to his adversely trained manner, which forces to produce a specific latent space distribution. In this way, all the features in the latent space are uncorrelated, *i.e.*, each feature captures a unique and specific characteristic of the input image

Conclusion

In this work, we examined the problem of detecting objects in luggage. The difficulty of the detection (inspection) includes a low prevalence, variation in the visibility, the search for an unknown set and overlapping objects. In this article, a contribution to the field of object recognition in X-ray testing for luggage control is presented by proposing a new deep learning architecture for semantic segmentation and object detection. The proposed architecture combines a deep convolution network with an Adversarial AutoEncoder (AAE) which acts as a useful feature extractor tool. The results obtained over the High Tech Detection Systems (HDTS) database show the superiority of the proposed system over five well-known semantic segmentation methods based on deep learning in order to distinguish the overlapping objects. The obtained results are very satisfactory with a small number of false detections, yielding the best performance in global accuracy, mean boundary F1 and mean IoU, with a percentage of 80.17%, 76.28%, and 76.84%, respectively, which is very encouraging for future works as the evaluation of other feature extraction alternatives and objects classification task. In perspective, future work will be devoted to the detection of threatening objects by means of double and multiple view X-ray images, as well as the detection of explosive objects by means of deep learning. And, implemented our systems in an electronic device.

References

Abidi BR, Zheng Y, Gribok AV, Abidi MA (2006) Improving weapon detection in single energy x-ray images through pseudocoloring. IEEE Transactions on Systems, Man, and Cybernetics 36(6):784–796

Agarap AF (2018) Deep learning using rectified linear units (relu). arXiv:1803.08375

Ba JL, Kiros JR, Hinton GE (2016) Layer normalization. arXiv:1607.06450

Babaheidarian P, Castañón D. (2017) Joint segmentation and material recognition in dual-energy ct images. International Symposium on Electronic Imaging, Computational Imaging XV 6(17):30–35

Badrinarayanan V, Kendall A, Cipolla R (2017) Segnet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Trans Pattern Anal Mach Intell 39(12):2481–2495

Chouai M, Merah M, Sancho-Gómez J-L, Mimi M (2019) Supervised feature learning by adversarial autoencoder approach for object classification in dual x-ray image of luggage. J Intell Manuf 31(1):1101–1112

DeDonato MP, Dimitrov V, Padır T (2014) Towards an automated checked baggage inspection system augmented with robots. In: Sensors, and Command, Control, Communications, and Intelligence (C3I) Technologies for Homeland Security and Homeland Defense XIII, volume 9074, Baltimore, Maryland, USA. International Society for Optics and Photonics

Gillen D, Morrison WG (2015) Aviation security: costing, pricing, finance and performance. Journal of Air Transport Management 48:1–12

Hassanpour R (2016) Illicit material detection using dual-energy x-ray images. International Arab Journal of Information Technology (IAJIT) 13(4):409–416

Heitz G, Chechik G (2010) Object separation in x-ray image sets. In: IEEE Computer society conference on computer vision and pattern recognition. IEEE, San Francisco, pp 2093–2100

Jiang J, Zheng L, Luo F, Zhang Z (2018) Rednet: Residual encoder-decoder network for indoor rgb-d semantic segmentation. arXiv:1806.01054

Liang J, Abidi BR, Abidi MA (2003) Automatic x-ray image segmentation for threat detection. In: Proceedings Fifth International Conference on Computational Intelligence and Multimedia Applications. ICCIMA 2003. IEEE, pp 396–401

Long J, Shelhamer E, Darrell T (2015) Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Boston, MA, USA, pp 3431–3440

Luigi d. S., Andrea B (1999) A simple and efficient connected components labeling algorithm. In: 10th International conference on image analysis and processing, Venice, Italy

Makhzani A, Shlens J, Jaitly N, Goodfellow I, Frey B (2015) Adversarial autoencoders. arXiv:1511.05644

Martin L, Tuysuzoglu A, Karl WC, Ishwar P (2015) Learning-based object identification and segmentation using dual-energy ct images for security. IEEE Trans Image Process 24(11):4069–4081

Megherbi N, Breckon TP, Flitton GT (2013) Investigating existing medical ct segmentation techniques within automated baggage and package inspection. In: Optics and photonics for counterterrorism, crime fighting and defence IX and optical materials and biomaterials in security and defence systems technology x, volume 8901, Dresden, Germany

Megherbi N, Flitton GT, Breckon TP (2010) A classifier based approach for the detection of potential threats in ct based baggage screening. In: IEEE 17th international conference on image processing, pp 1833–1836, Hong Kong, Hong Kong

Mehta S, Rastegari M, Caspi A, Shapiro L, Hajishirzi H (2018) Espnet: Efficient spatial pyramid of dilated convolutions for semantic segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), 552–568, Munich, Germany

Mery D (2014) Computer vision technology for x-ray testing. Insight - Non-Destructive Testing and Condition Monitoring (INSIGHT) 56(3):147–155

Mery D, Svec E, Arias M, Riffo V, Saavedra JM, Banerjee S (2017) Modern computer vision techniques for x-ray testing in baggage inspection. IEEE Transactions on Systems, Man, and Cybernetics: Systems 47(4):682–692

Michel S, Mendes M, Schwaninger A (2010) Can the difficulty level reached in computer-based training predict results in x-ray image interpretation tests? In: 44th Annual 2010 IEEE International Carnahan Conference on Security Technology. IEEE, pp 148–154

Muthukkumarasamy V, Blumenstein M, Jo J, Green S (2004) Intelligent illicit object detection system for enhanced aviation security. In: International Conference on Simulated Evolution and Learning (SEAL '04), Busan, Korea

Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E (2015) Deep learning applications and challenges in big data analytics. Journal of Big Data 2 (1)

Noh H, Hong S, Han B (2015) Learning deconvolution network for semantic segmentation. In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 1520–1528, Santiago, Chile

Peter VN (2015) Advanced integral passenger and baggage screening

Romera E, Alvarez JM, Bergasa LM, Arroyo R (2017) Erfnet: Efficient residual factorized convnet for real-time semantic segmentation. IEEE Trans Intell Transp Syst 19(1):263–272

Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: 18th international conference on medical image computing and computer-assisted intervention. Springer, Munich, pp 234–241

Rusu RB, Cousins S (2011) 3D is here: Point Cloud Library (PCL)

Sezgin M, Sankur B (2004) Survey over image thresholding techniques and quantitative performance evaluation. Electronic Imaging 13(1):146–166

Siam M, Gamal M, Abdel-Razek M, Yogamani S, Jagersand M, Zhang H (2018) A comparative study of real-time semantic segmentation for autonomous driving. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp 587–597, Salt Lake City, UT, USA

Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556

Sterchi Y, Schwaninger A (2015) A first simulation on optimizing eds for cabin baggage screening regarding throughput. In: 2015 International Carnahan conference on security technology (ICCST). IEEE, pp 55–60

Wiley D (2012) Analysis of anatomic regions delineated from image data. US Patent 8,194,964

Wiley DF, Ghosh D, Woodhouse C (2012) Automatic segmentation of ct scans of checked baggage. In: Proceedings of the 2nd International Meeting on Image Formation in X-ray CT, pp 310–313. Salt Lake City, USA

Zhao H, Qi X, Shen X, Shi J, Jia J (2018) Icnet for real-time semantic segmentation on high-resolution images. In: Proceedings of the European Conference on Computer Vision (ECCV), pp 405–420. Munich, Germany