A New Approach for Baggage Inspection by using Deep Convolutional Neural Networks

Ilhan AYDIN
Computer Engineering Department
Firat University
Elazig, Turkey
iaydin@firat.edu.tr

Mehmet KARAKOSE
Computer Engineering Department
Firat University
Elazig, Turkey
mkarakose@firat.edu.tr

Erhan AKIN
Computer Engineering Department
Firat University
Elazig, Turkey
eakin@firat.edu.tr

***Abstract*— In recent years, the use of x-ray equipment in different security point has increased. This equipment is heavily used at airports to control the baggage and bag of peoples. With this control, criminals are detected and terrorist acts can be prevented. This task is done by security officers at security points. It requires high concentration for the detection of threat objects. The manual operation of this process is both tedious and requires constant attention. There are many problems in the control with computer based automated systems. Because the position of the object in the baggage, overlapping with other objects makes the checking process difficult. In this study, a deep learning-based method for baggage control system was proposed by using x-ray images. The proposed method uses regions with convolutional neural networks for threat object detection. First, each objects in the images are labelled. Afterwards, the location of the image and bounding boxes of objects are given to regions with convolutional neural networks. The threat objects are detected and recognized in the last step. The proposed method is tested on a baggage inspection dataset and satisfied results are obtained.**

***Index Terms*— Deep learning, regions with convolutional neural networks, x-ray testing, threat object detection.**

I. INTRODUCTION

Security has become one of the most important issue in areas such as airports and shopping centers due to increased bombed or armed terrorist activities in recent years worldwide. For this purpose, x-ray devices are now being used in many shopping centers, municipalities and governorship buildings. The baggage and bag of people are examined on the computer screen by inspector people at security points whether people have threat objects in their bags. However, these objects are difficult to detect because they are placed too close together and occlusion.

Manually examining x-ray images by a human operator is a difficult and time-consuming process. At the same time, the human operator is receiving minimum technological support. Therefore, the development of computer-based algorithms facilitates the inspection process. Many methods have been proposed in the literature for extracting useful information from X-ray images. A fuzzy clustering based method is proposed for segmentation of dental images taken for medical diagnostic purposes [1]. The proposed method uses the neutrosophic cluster with fuzzy logic and the segmentation process gives better results than the other methods. X-ray imaging was used to classify the contents of shipping containers [2]. Image processing techniques such as SIFT and BOW have been applied to the images obtained for this purpose. The Naive Bayes method was used for classification purposes. In medical images, cervical vertebrae images have been segmented by using deep convolutional networks [3]. The results of segmentation can be used to detect cervical injuries that may not be noticed in emergency room. The x-ray images of a grain wheat have been classified by using principal component analysis and factor analysis [4]. Two analysis methods were compared and the results showed that the principal component analysis gives better results. The x-ray based inspection system was applied to fish fillets in order to detect fish bone [5]. The proposed method first segments the x-ray image and then extracts some density properties. These properties are evaluated by a classification method to classify boneless and boned fishes. In another study, a deep learning-based method for age determination was proposed for medical x-ray images [6]. The proposed method gives x-ray images to convolutional artificial neural network and age estimation is done. A GPU-based ray tracing method was applied to x-ray image in order to detect threat object in airline baggage inspection [7]. The proposed method used NVIDIA Optix API and the simulated results were verified by experimental results. A dictionary based method was proposed for threat object detection in x-ray images of a baggage [8]. The proposed method extracts some patches for each objects and constructs a dictionary for the related object. When a new test image arrives, the patches in this image are extracted and the type of object is determined according to the matching of the patches in the dictionary. Different computer vision methods have been proposed for baggage inspection from x-ray images. Different computer vision methods have been proposed for luggage control from x-ray images [9-10]. For this purpose, well-known computer vision algorithms such as bag of words, sparse representation and codebooks have been used to detect threat objects. The SIFT method is used for feature extraction.

978-1-5386-6878-8/18/$31.00 ©2018 IEEE

The x-ray imaging has begun to be widely used at airports and shopping centers for inspection as well as for medical purposes. Computer vision-based methods have been developed in the literature for automated inspection of x-ray images. However, the sensitivity of computer vision methods in different situations is low. At the same time, their modeling abilities are limited. In this study, a deep learning based method for baggage inspection was proposed. The proposed method can detect multiple objects using the Faster RCNN. The performance of the proposed method is higher than other computer vision based methods.

II. X-RAY IMAGING

Since X-ray imaging was discovered in 1985, this method is used to define internal structures of objects. This imaging method is used both for medical imaging system and for detecting the interior of materials and objects that cannot be detected with naked eye. Non-destructive testing with x-ray imaging is used in many applications such as inspection of crack of materials, quality control applications, baggage inspection, food quality analysis, cargo control, and quality control of electronic circuits [11]. A simple diagram of x-ray imaging system is given in Figure 1.

The two basic techniques used in X-ray imaging are digital radiography and computed tomography. Digital radiography uses digital radiographic films for imaging. On the other hand, digital tomography refers to electronic sensors to obtain the x-ray image of the target object [11]. Thanks to these features, digital tomography is faster than digital radiography and gives more accurate results. An amorphous silicon detector is used for the image sensor in digital tomography. X-ray imaging is modeled by absorption law. The basic principle is based on the characterization of density distributions of x-rays. The equation of absorption is given in (1).

$$\varphi(d) = \varphi e^{-\mu d} \tag{1}$$

In (1), μ and d represent absorption coefficient and thickness of a matter, respectively. The other two parameters φ_0 and φ show initial and normal energy flux densities. The gray value of an x-ray image can be given by a linear model as shown in (2).

$$I = \varphi . A + B \tag{2}$$

In (2), A and B are two constant parameters of a model. The model of an x-ray image can be used to simulate different conditions.

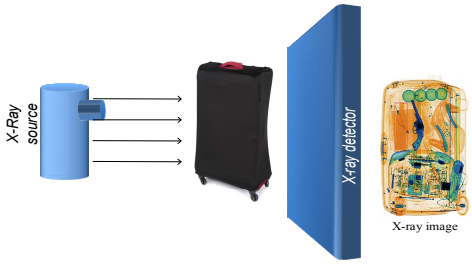


Fig.1. The block diagram of x-ray imaging

III. FASTER RCNN BASED OBJECT DETECTION AND RECOGNITION FROM X-RAY IMAGES

In this study, a deep learning based method is proposed for multiple object detection. The proposed method allows the detection and classification of specific objects within the x-ray image. The scheme of the proposed method is given in Figure 2.

In Fig.2, Faster RCNN has two main components: region proposal networks (RPN) and classifier [12]. The RPN is used for generating region proposal and the classifier component proposes potential objects. The faster RCNN uses a selective search algorithm for region suggestions [13]. After the potential regions have been obtained, the regressor is used to define the bounding box and the classifier is used to classify objects. The detailed components of the used in Faster RCNN is given in following sections.

A. Convolution neural networks

In convolutional neural networks, the most important mathematical process is the convolution process. This is accomplished by sliding a kernel matrix over the input image [14]. The kernel matrix slides on the input image and new values are calculated. A simple implementation of the convolutional operator is given in Figure 3.

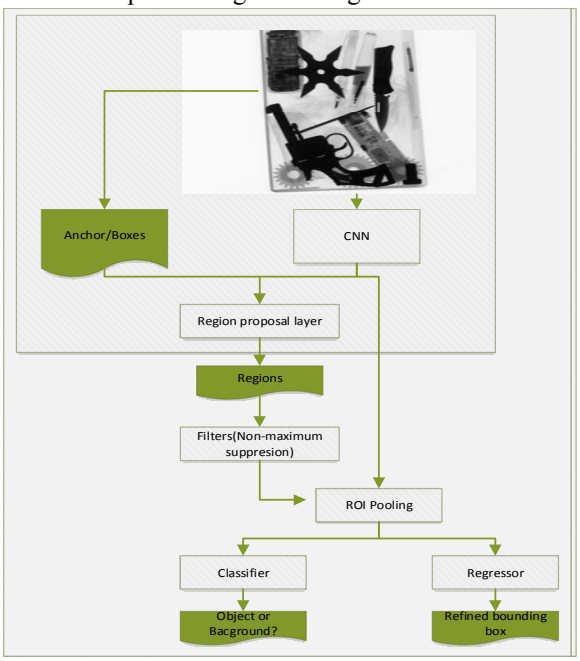


Fig.2. Faster RCNN based object detection for x-ray images

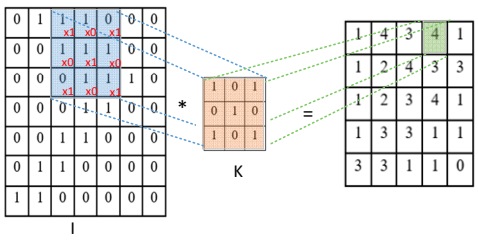


Fig.3. Convolution operator

As shown in Fig.3, a kernel function slides on an image and the mutual values of kernel matrix and related image part is multiplied and the new value is calculated with the sum of the multiplication operations. After the feature map has been constructed in the convolution layer, the anchors will be obtained from this map.

B. Anchors

Anchors are key components of faster RCNN. Each anchor is a box. At the last layer of a convolutional neural networks, a nxn sliding window is constructed. An anchor construction scheme is given in Figure 4.

Since anchors are obtained from a convolutional neural networks with the size of $conv_{height}xconv_{width}xconv_{depth}$, we use an image with size of $conv_{height}xconv_{width}$ to construct anchors. The size of sliding window is generally selected as 64, 128, or 256 pixels [12]. The ratio of size is selected as 0.5, 1, and 1.5. The value of p * should be calculated for each anchor. This parameter represents how much the anchor overlaps with the actual ground truth bounding boxes. This parameter is given in (3).

$$p^* = \begin{cases} 1, & if\ IoU > 0.7 \\ -1, & if\ IoU < 0.3 \\ 0 & otherwise \end{cases} \quad (3)$$

In (3), IoU represent the Intersection over Union and is given by (4).

$$IoU = \frac{Anchor \cap GTBox}{Anchor \cup GTBox} \quad (4)$$

In (4), GTBox represent the ground truth bounding box. After anchors have been obtained, they are given to region proposal layer.

C. Region Proposal Layer

Region proposal layer takes all anchors as input and obtains objects as output. RPN produces two different outputs for each anchor: probability of object and bounding box regression. RPN takes a 3x3 convolutional feature map as kernel with 512 features.

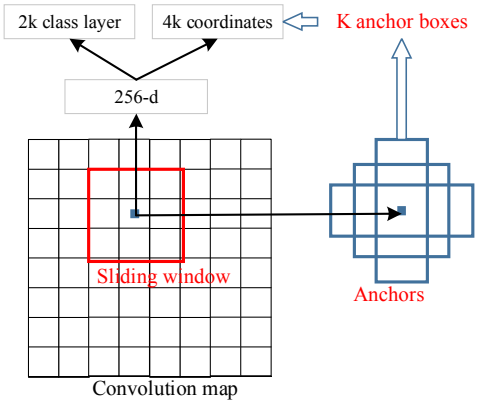


Fig. 4. Anchors at a sliding window

If there are k anchors available, 4k outputs are generated for regression and 2k output are generated for classification. In the classification, the given region is classified as object or background. In the regression, a bounding box is obtained which specifies the center points of object (x, y) and the width (w) and height (h) of the object.

There are many regions at the output of the region proposal layer. Some of these proposals overlap with another one. With non-maximum suppression, the number of proposal is reduced. This method takes a list of proposals sorted according to their scores and discards those proposals that have an IoU larger than a threshold value.

Region of Interest (ROI) pooling is used for object detection and improves the detection performance. As input, it takes a constant size feature map from convolutional neural networks and a matrix of NX5. The N value in the matrix indicates the number of RoIs. The first column shows the image index and the other four columns show the top and bottom coordinates of the region. After the inputs of RoI have been obtained, region proposals are divided into equal sized sections. Afterwards, maximum value of each section is found and copied these values to the output buffer. A simple example for RoI is given in Figure 5.

In Fig. 5, RoI is applied on a single 6x6 feature map and an output size 2x2. The max value of each section is copied to the output buffer.

D. Classification and Detection Layer

The last step of the Faster RCNN algorithm detects and recognize the detected objects and displays them in rectangular bounding boxes. After obtaining the convolutional feature map from the image, the object proposals are obtained by using RPN. Features are then extracted with the RoI stage. In the last stage, two fully connected convolutional neural networks are used for classification purposes. The main aim of this stage is to classify proposals into one of the classes and to adjust the bounding box for the proposals. An illustration of this stage is given in Figure 6.

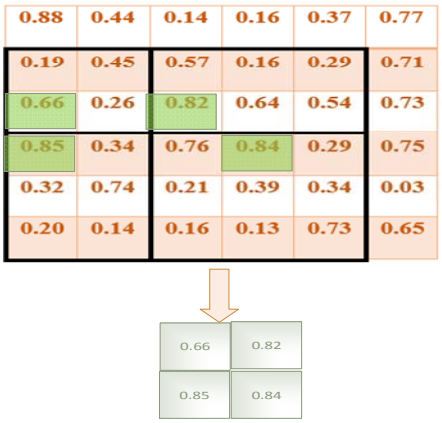


Fig. 5. RoI pooling

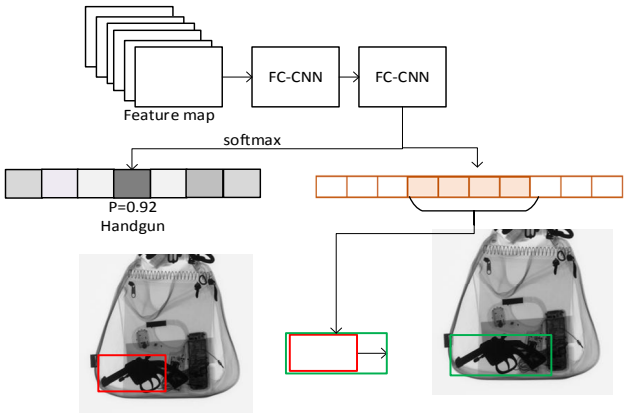
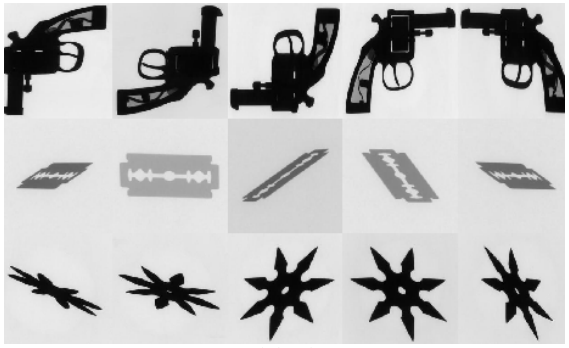


Fig. 6. Object detection and classification stages

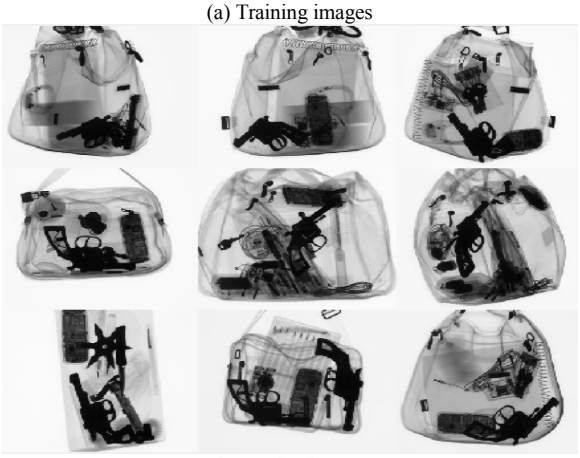
In Fig.6, we have two fully connected convolutional neural networks (FC-CNN). The first FC-CNN has n+1 class outputs and one of them is background. The other FC-CNN has 4N output and is used to determine the bounding boxes of objects with regression prediction.

IV. Application Results

Baggage x-ray images taken from the GDXray database were used in the study [15]. For this purpose, 2000 images were manually labeled. MATLAB ImageLabeler application is used for labeling the objects in the image. We have tried to detect handgun, razor and shuriken objects from X-ray images. Some training and testing images are given in Figure 7.



(a) Training images



(b) Testing images

Fig.7. Some of training and testing images

Figure 7 (a) shows the training data taken at different angles. The test phase presents these threat objects in a baggage. The Faster RCNN network consists of a total of 3 layers. The first layer is the image input layer and is the same size as the training images. In object detection, the input size is chosen larger than the size of the largest object, since the main aim of object detection is to find smaller areas in the image. The middle layer consists of convolution, RELU and maxpooling layers. The final layer has two components: fully connected convolutional neural networks and softmax layer. The layered structure and properties of the proposed Faster RCNN are given in Table I.

Training of Faster RCNN is completed in four stages. The first stage is the training of the region proposal network (RPN). In the second stage, training of Fast RCNN is done using RPN obtained in the first step. In the third step RPN, RPN is re-trained using weight sharing with Fast RCNN. In the last stage, Fast-RCNN is trained with updated RPN. The training of Faster RCNN was implemented on a laptop computer that has 2GB GTX940 GPU with 16GB RAM. Fig. 8 represents the training convergence of each step.

TABLE I. The layered structure of Faster RCNN

Layer	Layer name	Properties
1	Image input	32x32x1
2	Convolution	32 3x3 with stride [1 1] and padding [1 1 1 1]
3	RELU	-
4	Convolution	32 3x3 with stride [1 1] and padding [1 1 1 1]
5	RELU	-
6	Max pooling	3x3 max pooling with stride [2 2] and padding [0 0 0 0]
7	Fully connected	with 4 classes output
8	RELU	-
9	Fully connected	2 fully connected layer
10	Softmax	-
11	Classification output	-

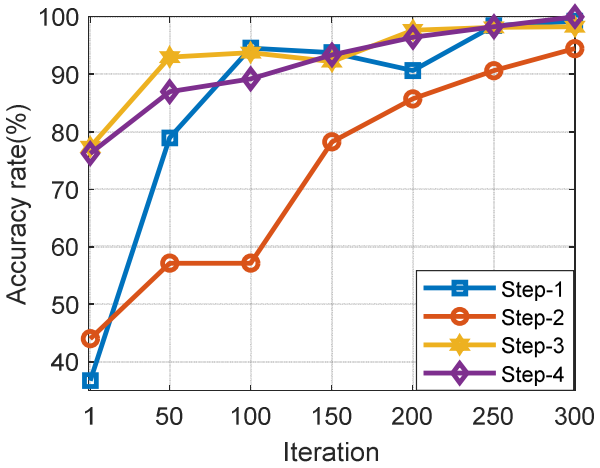


Fig. 8. The training accuracies of steps of Faster RCNN

In Fig. 8, each accuracy rate is obtained by operating each step with a given iteration number. The network has higher accuracy rates after re-training of RPN. Fig. 9 shows the detection results for some images.

As shown in Figure 9, different threat objects in the image have been detected. In the same image, two handguns were correctly detected. In addition, shuriken and handgun were detected as two different objects in the image. It is determined both object detection and what is the object. Therefore, both detection and recognition are performed in the image.

In order to determine the accuracy rate of the proposed method, accurate detection rates of objects are used in total images. A total of 1000 images are found in the dataset. In Table 2, the number of objects, true and false detection rates are given in 1000 images.

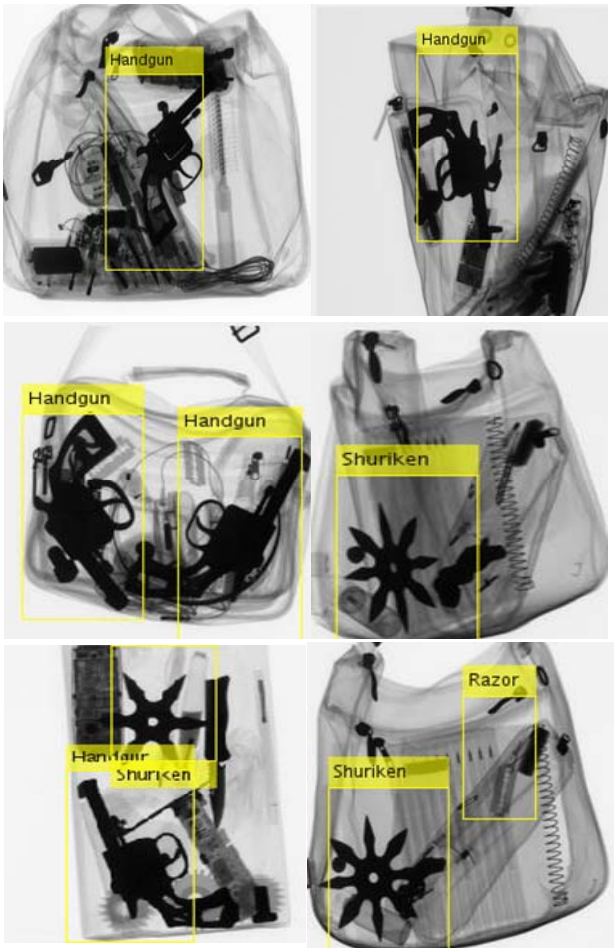


Fig. 9. Detection results of some objects

TABLE II. THE DETECTION PERFORMANCE OF THE PROPOSED METHOD

Object	Number of objects	True detection	False detection	Accuracy (%)
Handgun	537	525	12	97.77
Shuriken	414	410	4	99.13
Razor	368	362	6	98.37
Total				98.42

As shown in TABLE II, the detection rate of the proposed method is quite high. False detection or undetectable objects are objects in the bag that have a complex background. The proposed method was compared to the computer vision based methods in the literature. True detection rates have been used for comparison purposes. Figure 10 shows the comparison results.

In Figure 10, the proposed method was compared to five methods. The detection rate of the SURF method is very low. The detection performance is higher when the matching is done by using SIFT method. BoW creates a dictionary by taking patches from feature points obtained with SIFT. When any test object is given, the object is classified according to the dictionary matches. For each image in the sparse KNN method, the objects are segmented by the adaptive K-means method. Afterwards, SIFT descriptors are generated and related important features are selected. The AISM method generates a visual dictionary and descriptors by using key points. An object is defined as a visual dictionary of parts. Key points and their visual descriptors are generated by taking all training examples. At this stage, key points are identified with the SIFT method. In the test phase, the target objects are determined according to similarities. The main disadvantage of the computer vision based methods is that they are very sensitive to the image acquisition environment. In addition, the dictionary based method works quite slowly. Application results show that the proposed method is better than other computer vision based methods. In addition, the test phase of dictionary based methods is also quite slow.

V. CONCLUSIONS

In this study, we present a deep learning-based method for object detection and recognition from x-ray images. The proposed method detects and classifies objects such as gun shuriken and razor blades in the x-ray image. The proposed method uses Faster RCNN which is a well-known and popular deep learning method for object detection and classification. For this purpose, 1000 images were manually labeled and training and test data set were constructed. In some of the images there can be two or more objects from the same object as well as different objects. The results of the application show that the proposed method provides better results than traditional computer vision techniques.

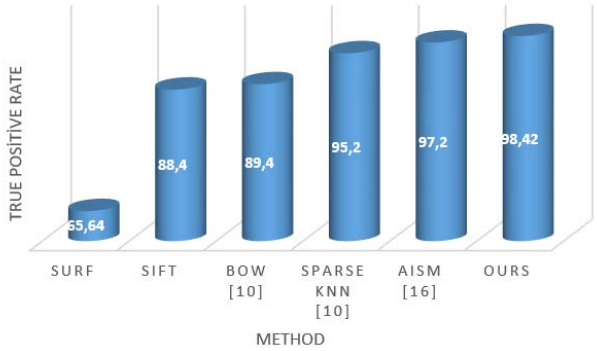


Fig.10. The comparison results with other methods

REFERENCES

[1] A. Mumtaz, M. Khan, and N. T.Tung, "Segmentation of dental X-ray images in medical imaging using neutrosophic orthogonal matrices," Expert Systems with Applications, vol. 91, pp. 434-441, 2018.

[2] A. Majid, M. Teimouri, and R. Rahmani. "Classification of X-Ray images of shipping containers," Expert Systems with Applications, vol. 77, pp. 57-65, 2017.

[3] S.M. M. Al Arif, K. Knapp, and G. Slabaugh, "Fully automatic cervical vertebrae segmentation framework for X-ray images," Computer Methods and Programs in Biomedicine, pp. 95-111, 2018.

[4] M. Charytanowicz, P. Kulczycki, P.A. Kowalski, S. Łukasik, R. Czabak-Garbacz, "An evaluation of utilizing geometric features for wheat grain classification using X-ray images," Computers and Electronics in Agriculture, vol. 144, pp. 260-268, 2018.

[5] D. Mery, I. Lillo, H. Loebel, V. Riffo, A. Soto, A. Cipriano, J. M. Aguilera, "Automated fish bone detection using X-ray imaging," Journal of Food Engineering, vol. 105, pp. 485-492, 2011.

[6] C. Spampinato, S. Palazzo, D. Giordano, M. Aldinucci, R. Leonardi, "Deep learning for automated skeletal bone age assessment in X-ray images," Medical image analysis, vol. 36, pp. 41-51, 2017.

[7] M.A. Haidekker, L.D. K. Morrison, A. Sharma, E. Burke, "Enhanced dynamic range x-ray imaging". Computers in biology and medicine, vol. 82, pp. 40-48, 2017.

[8] Q. Gong, R.I. Stoian, D.S. Coccarelli, J.A. Greenberg, E. Vera, M.E. Gehm, "Rapid simulation of X-ray transmission imaging for baggage inspection via GPU-based ray-tracing," Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms, vol. 415, pp. 100-109, 2018.

[9] D. Mery, A.K. Katsaggelos, "A Logarithmic X-Ray Imaging Model for Baggage Inspection: Simulation and Object Detection," In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 57-65, 2017.

[10] D. Mery, E. Svec, M. Arias, V. Riffo, J. M. Saavedra, S. Banerjee, "Modern computer vision techniques for x-ray testing in baggage inspection," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, pp. 682-692, 2017.

[11] D. Mery, "Computer Vision for X-Ray Testing," Switzerland: Springer International Publishing, 2015.

[12] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: towards real-time object detection with region proposal networks," IEEE transactions on pattern analysis and machine intelligence, vol. 39, pp. 1137-1149, 2017.

[13] J. R. Uijlings, K. E. Van De Sande, T. Gevers, A.W. Smeulders, "Selective search for object recognition," International journal of computer vision, vol. 104, pp. 154-171, 2013.

[14] M.A. Ponti, L.S.F. Ribeiro, T. S. Nazare, T. Bui, J. Collomosse, "Everything you wanted to know about Deep Learning for Computer Vision but were afraid to ask." In: 2017 30th SIBGRAPI Conference on Graphics, Patterns and Images Tutorials (SIBGRAPI-T), pp. 17-41. IEEE 2017.

[15] D. Mery, V. Riffo, U. Zscherpel, G. Mondragón, I. Lillo, I. Zuccar,H. Lobel, M. Carrasco, "GDXray: The database of X-ray images for nondestructive testing," Journal of Nondestructive Evaluation, 34.4:1-12, 2015.

[16] D. Mery, E. Svec, M. Arias, "Object recognition in X-ray testing using adaptive sparse representations," Journal of Nondestructive Evaluation, vol. 35, pp. 34- 45, 2016.