Material classification in X-ray images based on multi-scale CNN

Emil Benedykciuk¹ · Marcin Denkowski¹ · Krzysztof Dmitruk¹

Received: 5 June 2020 / Revised: 28 September 2020 / Accepted: 3 October 2020 / Published online: 6 February 2021
© The Author(s) 2021

Abstract

Security X-ray baggage scanners provide images based on the different levels of radiation absorption by different materials. Images captured by such scanners are inspected by a human operator, which can slow down the verification process. To speed up inspection time, computer vision and machine learning methods are increasingly being used. While object recognition has been the subject of a huge number of articles, the problem of material recognition in X-ray images still requires some work to achieve equivalent accuracy. This paper focuses on the problem of discrimination of materials into several classes, such as organic substances or metals, in images obtained from dual-energy X-ray security scanners. We propose a new multi-scale convolutional neural network (CNN) for predicting the material class, in which five different sizes of patches are implemented parallelly to balance the trade-off between the increase in the receptive field and the loss of detail. We analyze some regularization methods and activation functions and their impact on the effectiveness of our architecture. The results were compared with other popular CNN architectures and demonstrate the superiority of our solution.

Keywords X-ray imaging · Material classification · Per-pixel classification · Convolutional neural networks · Multi-scale CNN

1 Introduction

Object detection and identification in digital RGB images is a widely discussed problem in the literature. There are many projects containing ready-to-use modules with pre-learned neural networks that perform object recognition [1]. However, X-ray images are different from photographic images. An X-ray image of the contents of a package may help to determine if there is any dangerous object inside and avoid a possibly threatening situation. Unfortunately, recognition of objects after their shape is often insufficient in the case of X-ray images because there are many objects with an undefined shape or, to put it more precisely, those that can take a very different shape, such as liquids, powdery substances and fabrics. With only shape information, it is often impossible to identify objects that are obscured by other objects. Moreover, in X-ray scans, a reference to the matter from which these objects are made is helpful. The main problem

in discriminating the materials of a given object from only a single projection in an X-ray image is to determine its thickness, density and composition. Two very different materials (e.g., steel and water) can give identical readings on the X-ray detectors if they have different densities and/or thicknesses. For this reason, multi-energy techniques are used that allow for making such a distinction for a single material. Dual-energy X-ray imaging (DEXA) is one such a well-known technique [2] that requires two measurements at different energies. However, because raw X-ray images are not always easy to analyze and interpret, some image processing methods like object detection, a frequency resolution increase, or a pseudocoloring are being used [3,4]. Overall, the problem of material discrimination has not been well investigated by the computer vision community because, in the domain of luggage inspection, a significant part of the work is focused on object detection. We believe that a complete system for identifying and classifying objects in X-ray scans should primarily use information about the material and, secondly, information about shape.

In this paper, we propose a method that classifies materials in DEXA scans into six main types. Our aim is to employ a CNN approach for the entire feature extraction, representation and classification process. More specifically, we optimize the CNN structure and fine-tune convolutional

✉ Marcin Denkowski
marcin.denkowski@umcs.pl

Emil Benedykciuk
emil.benedykciuk@umcs.pl

¹ Department of Computer Science, Maria Curie-Skłodowska University, Lublin, Poland

Material class	Examples	Threads
Background		
Light organic	Wood, Fabric	Explosives C4, Trotyl
Heavy organic	Plastics, Glass	Drugs, Heroin, Cocaine
Non-organic	Aluminium, Salt	Gunpowder
Metals	Iron, Steel, Brass	Guns, Bullets
Heavy metals	Silver, Gold	Jewellery contraband
Not-penetrable	Lead	Threat hiding

Fig. 1 Material pseudocolors and its classes used widely in X-ray security scanners (color figure online)

fully connected and other layers for feature-to-classification pipeline within this problem domain. We perform experiments that illustrate the effect of various architectural decisions (i.e., regularization methods, number of layers and convolution filters) on the possibility of problem generalization. As a result of the proposed method, a per-pixel probability map will be created, mapping each point of the X-ray scan to one of the six classes of materials: background, light organic, heavy organic, light metals, heavy metals and non-penetrable, giving some form of initial segmentation. We believe that a complete system for identifying and classifying objects in X-ray scans should primarily use information about the material and, secondly, information about shape, especially in the case of contraband organic materials such as cigarettes, drugs, powders, explosives or liquids. Such materials do not have a specific shape so algorithms for their detection often fail.

This paper is organized as follows. Section 2 discusses the related works and summarizes the pros and cons of each of them. Section 3 gives all the necessary details of the proposed method. Section 4 then combines and analyzes the results from the experiments, while in Section 5 the results are compared to various popular CNN architectures designed to recognize visual patterns. Finally, Sect. 6 presents our conclusions and planned future works.

2 Related works

Dual-energy X-ray imaging requires two measurements at different energies providing two images based on different levels of radiation absorption of different materials. A common approach is to visualize these images using a linear color map (LCM). There are four or six main colors used widely in X-ray scanners to label material classes (see Fig. 1).

More advanced methods allow for classifying scanned objects or their parts with certain materials on the basis of a mass attenuation coefficient [20]. This coefficient depends on

the material's atomic number. In theory, it should allow us to classify the object on the basis of its atomic number, but our tests showed that an object's thickness has a major impact on the classification. As discussed in [26], the unambiguous definition of a material class for a composite of more than three substances is unachievable in the case of a fixed X-ray tube system. Due to this fact, several approximate hand-crafted methods have been proposed in [5–7] (see Table 1 for details). However, such classic material discrimination methods used for the dual-energy X-ray scans do not cope well with determining the type of material when there are many layers of different types of material at a given point. In the literature, several approaches have used classical machine learning methods in the automated inspection of X-ray images of airport baggage [9,10,14] and cargo [17,18], object/thread detection [15,16,16,18,27], sub-component level segmentation strategies for supervised anomaly detection [13,15,17], or material identification [12]. The use of deep learning techniques allows real-time and accurate detection of prohibited items even in cluttered X-ray images, although this is very often the case with already segmented or colored images. In our previous work, we examined several machine learning techniques, such as SVM or Random Forest, for material prediction in X-ray scans [11], where the obtained classification results can be used for initial image segmentation. It should be added that there are many works where machine learning methods have been used for material recognition in conventional images [21,22,28,29]. Authors in [19,30] showed that CNN also could be used for the segmentation of large materials obtained using X-ray computed tomography.

The presented literature, summarized in Table 1, shows that there are relatively few methods based on deep learning used to classify materials in X-ray images. This type of problem is more popular in the domain of traditional images, while most X-ray/dual(multi)-energy X-ray or CT solutions focus on the detection of objects, threats, anomalies, or image classification. But such algorithms do not work well for objects of undefined shape, such as powders or liquids, so the methods proposed in related works do not solve the problem.

3 Proposed method

3.1 Proposed CNN architecture

The X-ray images processed by the proposed classification system are two-channel images (low and high energy) with 16-bit integer precision. We want to classify the materials into six groups: background, light organic, heavy organic, light metals, heavy metals and impenetrable. A single training sample is a set of patches with different sizes for the specific type of material. The patches have the following

Table 1 A summary of the literature on X-ray security imaging in terms of the task and methods used

Reference	Task	Methods	Notes
Hand-crafted methods, X-ray images			
Alvarez et al. [5]	Approximate atomic number Z, material classification	Direct approximation of polynomials	Approximate atomic number for only two overlapping materials
Chuang et al. [6]	Approximate atomic number Z, material classification	Linear ISO-transmission, subregion direct approximation methods	Methods are computationally more efficient than the nonlinear equations method, but more sensitive to noise
Chen et al. [7]	Material recognition	Curved-based, smooth strategy, HLS color space image colorization processing	Proves the calibration strategy of the classification curve enhances the classification performance
Osipov et al. [8]	Approximate atomic number Z, material classification	Statistical processing of radiographic images	Developed to compensate the effect of container wall thickness on identification, classes of materials: light organics, mineral materials, light metals, calcium, metals, heavy metals
Roomi et al. [9]	Object detection, object classification (threat detection)	Fuzzy k-NN classifier with a shape context descriptor and Zernike moments	Limited evaluation over only 15 image examples
Kundegorski et al. [10]	Image classification, (threat detection, firearms detection)	BoVW model with FAST-SURF feature detector and descriptor and SVM and RF classification	Bag-of-Visual-Words (BoVW)-based approaches for threat object detection, an optimal statistical accuracy of 0.94 is achieved
Machine learning-based methods, X-ray images			
Benedykciuk et al. [11]	Material classification	Random Forest, SVM	Patch material classification, average probability of the class for each pixels based on all patch sizes
Brumbaugh et al. [12]	X-ray Diffraction Tomography; material identification	1D CNN	Use chemical/molecular signatures that can be used to identify a target material; trained on simulated data
Bhowmik et al. [13]	Object and sub-component level segmentation, image classification (anomaly, threat detection)	Faster R-CNN, Simple Linear Iterative Clustering (SLIC), CNN (VGG-16, SqueezeNet, ResNet), Fine-grain classification (BCNN, MA, DFL)	Sub-component level segmentation approach in combination with a specific fine-grain CNN architecture achieving a performance accuracy of 97.91% with a notable 3.50% false-positive rate for realistic anomaly concealment within representative consumer electronic items
Akay et al. [14]	Object classification, (handgun detection)	CNN (transfer learning)	Accuracy of handgun detection 98.92%
Akay et al. [15]	Object detection, classification and segmentation	Classical machine learning approaches, deep learning algorithms	Paper taxonomizes conventional machine and modern deep algorithms utilized within X-ray security imaging
Gaus et al. [16]	Object detection	CNN (Faster R-CNN, Mask R-CNN, RetinaNet)	Faster R-CNN achieves superior performance on 91% firearm and 88% firearm components; use two datasets: Dbf2 (Durham Dataset Full Two-class), SIXray (public dataset)
Andrews et al. [17]	Image classification (anomaly, threat detection)	One-class RBF SVM classifiers feeding by feature vectors	Feature vectors derived from a sparse autoencoder feed-forward neural network
Jaccard et al. [18]	Image classification, threat detection	CNN 18-layer (FC1) + input images were log-transformed prior to features computation	100% car image classification rate for a false-positive rate of 1-in-454

Table 1 continued

Reference	Task	Methods	Notes
Computed tomography (CT)			
Stan et al. [19]	Material semantic segmentation	CNN autoencoder (Seg-Net)	The segmented images are quantitatively superior to what could be obtained using conventional segmentation techniques
Flitton et al. [20]	3D object and artifacts detection	Local region histograms, 3D RIFT and SIFT descriptors	Recognition rates in excess of 95% are demonstrated with minimal false-positive rates for a set of exemplar 3D objects
Machine learning-based methods, common images			
Bunrit et al. [21]	Image (material) classification	CNN (transfer learning)	Classification of construction material images, fine-tuning scheme of GoogleNet reveals the highest classification result by 95.50% of accuracy
Roy et al. [22]	Image (material) segmentation	Multi-scale CNN	Three multi-scale CNNs predict mid-level cues, combined by another multi-scale CNN for predicting dense affordance maps of a single RGB image
ImageNet [23–25]	Object detection and classification	CNN (transfer learning)	Networks trained using all patch sizes from the MDD dataset [11], accuracy results comes from ensemble networks

sizes: 3×3 , 5×5 , 7×7 , 9×9 , 15×15 . As attributes of materials are not defined semantically, we annotate every set of training patches with the appropriate label.

It is important to note that the statistical properties of different patch sizes of the input data can vary largely, which makes it difficult for a single, sequential model to directly encode such data (i.e., by simply concatenating the data and then applying the single-channel model). To overcome such a difficulty, we need a multi-scale model that gives us better capability of modeling multi-channel, multi-scale input data and fuses them together to generate high-level features. Inspired by deep convolutional networks (CNN) [25] and multi-scale architectures [22,30] we propose our version of multi-scale network with five inputs fed with different patch sizes for giving a final material class on the output.

Figure 2 shows the schema of the proposed multi-scale convolutional neural network that consists of five subnetworks, each with a different structure depending on the size of the input patch. As presented in [25], to increase the performance of our convolutional neural network, we adapted our model subnetworks to the resolution of received patches. As the patch resolution increases, the subnet goes deeper and wider. The output of each subnetwork is a feature vector. The feature vectors from all five subnetworks are concatenated and passed to the serialized, two fully connected layers (FC) finalized by the softmax layer. This enables training CNNs based on multiple input scales.

Specifically, we regard the outputs from the last two layers of the CNN as the learned high-level appearance features of multiple input patches. It is essential to extract image features in a precise way. Despite the fact that the architecture of our CNN is not very complex, it allows for a good extraction

Fig. 2 Our approach multi-input CNN schema (color figure online)

of features. This is the result of a hierarchical arrangement of successive layers of the subnetwork. What's more, we used an exponent-linear unit (ELU) as the activation function proposed in [31]. Their work proved that the result of the ELU was better than all the varieties of the rectified linear units (ReLU) function, resulting in shorter learning time and better neural network performance against the test set.

3.2 Training details

Our CNN classifier has been trained on input data based on low (LE) and high (HE) energy X-ray readings. Input data consisting of two X-ray energies compose the following three-channel image: (1) HE, (2) LE and (3) filled with zeros.

We train our CNN by fine-tuning the network, starting from the weights initiated using the Xavier initialization strategy. During training, we use an adaptive momentum estimation (Adam) optimizer with a 2048 batch size and a constant learning rate of $1e-4$ (decay is zero). The choice of the optimization of the cost function was motivated by the fact that we trained the classifier to predict only one class (i.e., a multi-class model, not multi-output), so we use a softmax function (also called a normalized exponential function). The purpose of training is to obtain a model that estimates a high probability for the target class and at the same time a low probability for the other classes. Therefore, we use cross-entropy [32] as the cost function.

In order to avoid overfitting our model, we examined various regularization methods. One of the most important is a dropout layer with 0.5 rate (during the training process) proposed by Hinton et al. [33]. The accuracy metric was used to estimate network efficiency and its ability to generalize during the processing of validation data. The learning process was run 10,000 times for the entire training dataset (10,000 epochs). However, for many epochs, an early stopping regularization method is also used when the model has not improved for 100 epochs. In addition, we explored regularization to all convolutional layers. Regularizers allow for applying penalties on layer parameters or layer activity during optimization. These penalties are incorporated in the loss function that the network optimizes. We chose activation regularization with the norms $|L_1| = 0.001$ and $|L_2| = 0.01$. Finally, we verified the difference in the effectiveness and generalization capabilities of our network for the Dropout layer and multiple DropBlock layers. DropBlock, introduced by Ghiasi et al. [34], is a form of structured dropout, where units in a contiguous region of a feature map are dropped together. As the authors proved, DropBlock works better than dropout in regularizing convolutional networks.

4 Experiments and results

The main parameters we have analyzed are the accuracy of material recognition efficiency in terms of learning speed and model size. The proposed model is implemented using Tensorflow and Keras libraries. All of our experiments are conducted on the Nvidia graphics processing unit GeForce 2080 (Turing microarchitecture).

All training and test data come from our “Materials in DEXA Scans Database” (MDD) presented in [11]. We

trained the classifier on a dataset comprised of over 1 million sample patches and over 100k test patches. Materials were classified into five groups: background, light organic, heavy organic, light metals and heavy metals. The last sixth class, i.e., impenetrable materials, was not trained as a degenerated class. All models created were trained and adjusted to the validation set, and the final results were made for the test set to verify the stability of the method.

4.1 Proposed multi-scale solution

Our estimator (without any regularization methods) quickly reaches the accuracy over 99% for the validation set (just after the 142nd epoch), which is very satisfactory. But if we look at the confusion matrix for the material classes in Tables 2(1) and 4, it turns out that the network itself is poor at distinguishing the classes of light organics, heavy organics and light metals. It allows us to conclude that, along with the subsequent epoch of the learning process, the model is overfitted. For this reason, we have tested various regularization methods that will allow for better generalization of material classification.

Analyzing Table 3, we can note the following: using the ELU activation function for the network we proposed has resulted in improved performance. In addition, we also see this increase in the accuracy of combining the ELU activation function with the regularization methods, i.e., L1, L2 and DropBlock, with the exception of the Dropout method. Another point to notice is the reduction of learning time for DropBlock, Dropout and L1 regularization methods. The large generalization possibilities provided by the Dropout, L1, L2 methods and the benefits of using the ELU activation function prompted us to verify the combination of all these elements, which is presented in the last two rows of Table 3. As can be seen, the combination of these elements brought the greatest prediction stability in both cases for validation and test datasets—the difference in accuracy between the validation and test datasets is much smaller compared to the other options presented in the table. The problem that appears is the number of epochs needed to train such a stable classifier. However, as will be shown in Sect. 5 and Table 5, training time will be much shorter compared to popular architectures.

In order to verify the best regularization methods for our network generalization predictions, we have prepared the confusion matrices in Table 2 for a test dataset with ELU activation function. We achieved the best result for material classification for the ELU activation function with the L1 and L1 + Dropout regularization methods composition. It can be also seen in Table 2 that all tested cases without the L1 or L2 regularization method have a significant problem with the classification of the heavy organic materials. This is the class of materials that intermingles with other neighboring classes the most, i.e., with light organics and

Table 2 Normalized confusion matrices for the validation and test dataset with ELU function and various regularization methods

	Validation					Test				
	0	1	2	3	4	0	1	2	3	4
(1) Baseline										
0	1.00					0.98				
1		1.00					1.00			
2			1.00				1.00			
3			0.01	0.99				0.11	0.89	
4					1.00					0.96
(2) Drop Block										
0	1.00					1.00				
1		1.00					1.00			
2			1.00				1.00			
3			0.01	0.99				0.26	0.74	
4					1.00					1.00
(3) Dropout										
0	1.00					1.00				
1		1.00					1.00			
2			1.00				1.00			
3			0.02	0.98				0.25	0.75	
4					1.00					1.00
(4) L1										
0	1.00					1.00				
1		0.99	0.01				0.97	0.02		
2		0.03	0.94	0.03			0.20	0.80		
3		0.01	0.11	0.86	0.03			0.19	0.81	
4					1.00			0.08		0.92
(5) L2										
0	1.00					0.98	0.01			0.01
1		0.98	0.02				1.00			
2		0.02	0.97	0.01			0.24	0.76		
3			0.11	0.88	0.01			0.49	0.51	
4				0.01	0.99			0.08		0.92
(6) L1 + Dropout										
0	1.00					0.97	0.01		0.01	
1		0.99	0.01				0.96	0.04		
2		0.01	0.96	0.03			0.24	0.76		
3			0.09	0.82	0.08			0.06	0.94	
4				0.01	0.99			0.02		0.98
(7) L2 + Dropout										
0	1.00					0.99	0.01			
1		0.99	0.01				1.00			
2		0.06	0.93				0.31	0.69		
3			0.18	0.81	0.01			0.23	0.77	
4				0.02	0.98			0.05		0.95
(8) Inception-ResNet-v2										
0						1.00				
1							1.00			

Table 2 continued

	Validation					Test				
	0	1	2	3	4	0	1	2	3	4
2							0.98	0.02		
3								0.14	0.86	
4									0.05	0.95

Material classes: 0—background, 1—light organics, 2—heavy organic, 3—light metals, 4—heavy metals. Predicted labels are on the horizontal axis, true labels are on the vertical axis. Only non-zero values are shown. The highest probability for predictions are shown in bold

light metals, which we also noticed during the verification of machine learning methods in [11]. However, as shown in Tables 3 and 2, it should be noted that not all regularization methods have resulted in increased neural network accuracy and predictability generalization.

In summary, the best results for our neural network were achieved by using the L1 + Dropout method with the average accuracy equal to 0.955. The DropBlock layer can be a bit disappointing, especially in relation to the results presented in [34], probably due to the high noise of the input data and the relatively small input images (patch sizes: 3×3 – 15×15). Thus, we received the model that obtains the highest accuracy and has the best opportunity to generalize the problem of material classification in DEXA scans.

5 Discussion

Due to the lack of multi-input and multi-scale convolutional network architectures that would allow full use of the capabilities of our datasets [11], we decided to compare our results to three ImageNet challenge (ILSVCR) winning architectures: (1) VGG16 [23], (2) InceptionResNet V2 [24] and (3) EfficientNet in B0 version [25] as well as for our previous solution based on Random Forest classifier [11]. A full comparison study including all the methods selected in Table 1 is, however, beyond the scope of this work.

In each of the ImageNet architectures, we only change the last three dense layers to the following number of units: 128, 64 and the output with 5 units (number of material classes). For research purposes, we checked the accuracy of neural networks for the default input sizes: 224×224 (VGG16, EfficientNet) and 299×299 (InceptionResNet V2). These resolutions are much larger than those provided by the MDD dataset. Prior to training, each patch was scaled with cubic interpolation to the appropriate size for the given architecture and labeled with the appropriate material class. All values in patches were scaled to the range [0..255] and composed into three-channel images from HE, LE and zeros. All the ImageNet networks selected were trained for each patch type

Table 3 The average accuracy of classification of all material classes for validation and test datasets for our multi-input CNN with various regularization methods

Activation function	Epoch	Regularization method	10				50				100				1000				Last epoch				Epoch number
			Accuracy																Accuracy				
			val	test	val	test	val	test	val	test	val	test	val	test	val	test	val	test					
ReLU	–		0.333	0.263	0.854	0.946	0.981	0.979	0.999	0.924	0.924	0.999	0.924	0.999	0.924	0.999	0.924	0.999	0.924	1005			
ReLU	dropblock		0.335	0.237	0.850	0.943	0.956	0.932	0.999	0.894	0.894	1.000	0.896	1.000	0.896	1.000	0.896	1.000	0.896	1152			
ReLU	dropout		0.314	0.237	0.814	0.936	0.941	1.000	–	–	–	0.999	0.971	0.999	0.971	0.999	0.971	0.999	0.971	988			
ReLU	L1		0.173	0.456	0.418	0.282	0.418	0.282	0.883	0.813	0.813	0.930	0.800	0.930	0.800	0.930	0.800	0.930	0.800	6384			
ReLU	L2		0.319	0.237	0.431	0.282	0.438	0.282	0.974	0.866	0.866	0.977	0.877	0.977	0.877	0.977	0.877	0.977	0.877	1101			
ELU	–		0.379	0.248	0.915	0.933	0.991	0.927	0.999	0.943	0.943	1.000	0.928	1.000	0.928	1.000	0.928	1.000	0.928	1056			
ELU	dropblock		0.374	0.240	0.920	0.991	0.979	0.891	–	–	–	0.999	0.919	0.999	0.919	0.999	0.919	0.999	0.919	785			
ELU	dropout		0.333	0.237	0.853	0.953	0.990	0.948	–	–	–	1.000	0.924	1.000	0.924	1.000	0.924	1.000	0.924	799			
ELU	L1		0.313	0.237	0.417	0.247	0.338	0.237	0.808	0.922	0.922	0.960	0.944	0.960	0.944	0.960	0.944	0.960	0.944	5805			
ELU	L2		0.316	0.237	0.527	0.393	0.491	0.742	0.922	0.964	0.964	0.967	0.903	0.967	0.903	0.967	0.903	0.967	0.903	5442			
ELU	L1+dropout		0.371	0.237	0.426	0.282	0.396	0.238	0.811	0.920	0.920	0.956	0.955	0.956	0.955	0.956	0.955	0.956	0.955	5904			
ELU	L2+dropout		0.150	0.455	0.430	0.277	0.500	0.289	0.917	0.959	0.959	0.929	0.920	0.929	0.920	0.929	0.920	0.929	0.920	5930			

In each column for the last epoch, the top three scores are in bold

and size, and their accuracy was verified based on ensemble predictions for each patch size from a given perceptual area.

The results of the accuracy of these architectures and our proposal are presented in Table 5. The results are the ensemble accuracy of all patch sizes for classification of all material classes and are calculated based on the class prediction that obtained the highest probability in total of all patch sizes. We chose the ensemble accuracy for ImageNet architectures because it more closely resembles the operation of the model we proposed. The time needed to train the model in Table 5 is only a rough comparison, but it can be seen that the solution we propose is much faster and much more effective—in particular after adding regularization methods. As we can see in Table 2(8), in the confusion matrix the InceptionResNet-v2 (the best of the compared ImageNet models) incorrectly classifies materials from the heavy organics and light metal classes.

The weakest point in our comparison is the initial interpolation of the input data, which is necessary due to the architecture of the investigated ImageNet networks. Unfortunately, images from X-ray scanners are often very noisy. Interpolation of such images will cause this noise to be blurred and enlarged. This leads to recognizing the noise as a feature of a given material class, not artifacts that should be omitted during the process of learning material attributes. And as shown in [35], this type of quality distortions of an image affect the effectiveness of solutions based on convolutional neural networks. The networks are more sensitive to changes in blur and noise compared with compression and contrast. Most likely the above and the fact that the models are too complex for the problem presented by us makes the results of our solution much better. In particular, this can be seen on the confusion matrix (presented in Table 2), which depicts well the problem with the interpenetration of the classes of materials. Additionally, the network we proposed learns much faster, which is related to the fact that there are fewer learning parameters.

6 Conclusion and future scope

We successfully developed a multi-scale convolutional network architecture that classifies materials with per-pixel precision in DEXA scans into six main types with very high accuracy and achieve better results in comparison to popular CNN architectures and a machine learning method. The presented method creates a per-pixel probability map of belonging to the appropriate material class and can only be treated as an initial segmentation. We also confirmed our assumptions from [11] that deep learning methods will achieve better results in this problem than machine learning algorithms. In our publication, we analyzed some regularization methods and their impact on the effective-

Table 4 Classification metrics for test dataset for the proposed CNN with ELU activation function and L1+Dropout regularization methods

Class	Precision		Recall		F1-score		Support	
	val	test	val	test	val	test	val	test
0	0.99	1.00	1.00	0.97	1.00	0.99	340	556
1	0.99	0.94	0.99	0.96	0.99	0.95	731	289
2	0.92	0.65	0.96	0.76	0.94	0.70	540	55
3	0.94	0.94	0.82	0.94	0.88	0.94	396	135
4	0.90	1.00	0.99	0.98	0.94	0.99	311	184
Avg	0.95	0.96	0.95	0.96	0.95	0.96	2318	1219

Material classes: 0—background, 1—light organic, 2—heavy organic, 3—light metals, 4—heavy metals

Table 5 Comparison of the estimated training time, the total number of parameters and an average accuracy of the classification all material classes for a test dataset for all considered architectures

CNN Architecture	Train time [min]	Total parameters	Accuracy
VGG16	~ 333	17 934 661	0.888
InceptionResNet-v2	~ 842	66 867 813	0.892
EfficientNet-B0	~ 766	12 044 417	0.860
Random Forest [11]	~ 3	—	0.935
Our architecture	~ 58	347 109	0.955

The results for our architecture are shown in bold

ness of our architecture. Additionally, by analyzing the two activation functions ReLU and ELU, we have shown that also for the problem presented by us and for the multi-scale neural network, the function of ELU activation brings benefits in the form of better accuracy and a faster learning process. A considerable advantage, in terms of the accuracy and speed, was observed for the proposed network-based method, with an accuracy improvement of approximately 7.30% compared to the best of selected ImageNet architectures (InceptionResNet-v2) and with 2.14% compared to our previous solution based on Random Forest classifier.

There are still some problems with the classification of the heavy organic material, which intermingles with other neighboring classes, i.e., with light organics and light metals, but the same problem was also noticed during the verification of other discussed methods. The limitation of the proposed method is also the classification into six fixed classes of materials, which, however, is motivated by the use of the method in typical security scanners.

Our next goal is to develop a method that allows for more precise material discrimination on full resolution scans and performs image segmentation, which will contribute to the development of this field of science and increase the efficiency of solutions for more advanced computer vision problems in X-ray images, i.e., object detection. There remains the problem of noisy X-ray images, which probably have the greatest impact on the classification and segmentation of such images. In order to solve this obstacle, we want to try out deep neural networks (autoencoders), whose aim is

to denoise images or simultaneously interpolate images with removing noise.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Jiao, L., Zhang, F., Liu, F., Yang, S., Li, L., Feng, Z., Qu, R.: A survey of deep learning-based object detection. *IEEE Access* **PP**, 1 (2019)
- Perisinakis, K.: Dual-energy x-ray computed tomography. In: Russo, P. (ed.) *Handbook of X-Ray Imaging*, Ch. 39. CRC Press, Boca Raton (2018)

3. Dmitruk, K., Denkowski, M., Mazur, M., Mikołajczak, P.: Sharpening filter for false color imaging of dual-energy x-ray scans. *Signal Image Video Process.* **11**(4), 613–620 (2018)
4. Dmitruk, K., Mazur, M., Denkowski, M., Mikołajczak, P.: Method for filling and sharpening false colour layers of dual energy x-ray images. *Int. J. Electron. Telecommun.* **62**(1), 49–54 (2016)
5. Alvarez, R., Macovski, A.: Energy-selective reconstructions in x-ray computerized tomography. *Phys. Med. Biol.* **21**, 733–744 (1976)
6. Chuang, K.-S., Huang, H.K.: Comparison of four dual energy image decomposition methods. *Phys. Med. Biol.* **33**(4), 455–466 (1988)
7. Chen, Z.-Q., Zhao, T., Li, L.: A curve-based material recognition method in MeV dual-energy x-ray imaging system. *Nucl. Sci. Tech.* **27**, 11 (2014)
8. Osipov, S., Usachev, E., Chakhlov, S., Shchetinkin, S., Song, S., Zhang, G., Batranin, A., Osipov, O.: Limit capabilities of identifying materials by high dual- and multi-energy methods. *Rus. J. Nondestr. Test.* **55**, 687–699 (2019)
9. Roomi, M.: Detection of concealed weapons in x-ray images using fuzzy k-NN. *Int. J. Comput. Sci. Eng. Inf. Technol.* **2**, 187–196 (2012)
10. Kundegorski, M., Akçay, S., Devereux, M., Mouton, A., Breckon, T.: On using feature descriptors as visual words for object detection within x-ray baggage security screening. In: 7th International Conference on Imaging for Crime Detection and Prevention (ICDP 2016), pp. 12 (6.)–12 (6.), 01 (2016)
11. Benedykciuk, E., Denkowski, M., Dmitruk, K.: Learning-based material classification in x-ray security images. In: Farinella, G.M., Radeva, P., Braz, J. (eds.), *Proceedings of the 15th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, Volume 4: VISAPP, Valletta, Malta, 27–29 February, 2020*, vol. 4, pp. 284–291, SciTePress, 02 (2020)
12. Diallo, S., Gregory, C., Royse, C., Greenberg, J., Roe, K., Brumbaugh, K.: Material classification using convolution neural network (CNN) for x-ray based coded aperture diffraction system. In: *Conference Presentation*, p. 10, 05 (2019)
13. Bhowmik, N., Gaus, Y. F. A., Akçay, S., Barker, J. W., Breckon, T. P.: On the impact of object and sub-component level segmentation strategies for supervised anomaly detection within x-ray security imagery. In: Wani, M. A., Khoshgoftaar, T. M., Wang, D., Wang, H., Seliya, N. (eds.), *ICMLA*, pp. 986–991, IEEE, (2019)
14. Akçay, S., Kundegorski, M. E., Devereux, M., Breckon, T. P.: Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery. In: *2016 IEEE International Conference on Image Processing (ICIP)*, pp. 1057–1061 (2016)
15. Akçay, S., Breckon, T. P.: Towards automatic threat detection: a survey of advances of deep learning within x-ray security imaging. [arXiv:2001.01293](https://arxiv.org/abs/2001.01293) (2020)
16. Gaus, Y. F. A., Bhowmik, N., Breckon, T. P.: On the use of deep learning for the detection of firearms in x-ray baggage security imagery. In: *2019 IEEE International Symposium on Technologies for Homeland Security (HST)*, pp. 1–7 (2019)
17. Andrews, J., Morton, E., Griffin, L.: Detecting anomalous data using auto-encoders. *Int. J. Mach. Learn. Comput.* **6**, 21 (2016)
18. Jaccard, N., Rogers, T., Morton, E., Griffin, L.: Detection of concealed cars in complex cargo x-ray imagery using deep learning. *J. X-Ray Sci. Technol.* **25**, 06 (2017)
19. Stan, T., Thompson, Z., Voorhees, P.: Optimizing convolutional neural networks to perform semantic segmentation on large materials imaging datasets: X-ray tomography and serial sectioning. *Mater. Charact.* **160**, 110119 (2020)
20. Flitton, G., Breckon, T., Megherbi, N.: A comparison of 3d interest point descriptors with application to airport baggage object detection in complex CT imagery. *Pattern Recogn.* **46**, 2420–2436 (2013)
21. Bunrit, S., Kerdprasop, N., Kerdprasop, K.: Evaluating on the transfer learning of CNN architectures to a construction material image classification tasks. *Int. J. Mach. Learn. Comput.* **9**, 201–207 (2019)
22. Roy, A., Todorovic, S.: A multi-scale CNN for affordance segmentation in RGB images. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) *ECCV (4). Lecture Notes in Computer Science*, vol. 9908, pp. 186–201. Springer, Cham (2016)
23. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. [arxiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
24. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A. A.: Inception-v4, Inception-ResNet and the impact of residual connections on learning. In: *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI '17*, pp. 4278–4284, AAAI Press (2017)
25. Tan, M., Le, Q. V.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.), *Proceedings of Machine Learning Research, ICML*, vol. 97, pp. 6105–6114, PMLR (2019)
26. Rebuffel, V., Dinten, J.-M.: Dual-energy x-ray imaging: benefits and limits. *Insight Non-Destr. Test. Cond. Monit.* **49**, 589–594 (2007)
27. Petrozziello, A., Jordanov, I.: Automated deep learning for threat detection in luggage from X-Ray images. pp. 505–512. 11 (2019)
28. Bian, P., Li, W., Jin, Y., Zhi, R.: Ensemble feature learning for material recognition with convolutional neural networks. *EURASIP J. Image Video Process.* **2018**, 12 (2018)
29. Xu, S., Muselet, D., Treméau, A.: Deep learning for material recognition: most recent advances and open challenges. In: *Proceedings of the International Conference on Big Data, Machine Learning and Applications (BIGDML) At: Silchar*, vol. 10, India (2019)
30. Liu, X., Hou, F., Qin, H., Hao, A.: Multi-view multi-scale CNNs for lung nodule type classification from CT images. *Pattern Recognit.* **77**, 262–275 (2018)
31. Clevert, D.-A., Unterthiner, T., Hochreiter, S.: Fast and accurate deep network learning by exponential linear units (elus). In: Bengio, Y., LeCun, Y. (eds.), *ICLR (Poster)* (2016)
32. Goodfellow, I., Bengio, Y., Courville, A.: *Deep Learning*. The MIT Press, Cambridge (2016)
33. Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors. [arXiv:1207.0580](https://arxiv.org/abs/1207.0580) (2012)
34. Ghiasi, G., Lin, T.-Y., Le, Q. V.: Dropblock: a regularization method for convolutional networks. In: *Proceedings of the 32nd International Conference on Neural Information Processing Systems, NIPS'18, (Red Hook, NY, USA)*, pp. 10750–10760, Curran Associates Inc. (2018)
35. Dodge, S. F., Karam, L. J.: Understanding how image quality affects deep neural networks. In: *QoMEX*, pp. 1–6, IEEE (2016)