Steffen Rüger*, Markus Firsching, Julija Lucic, Alexander Ennen, Norman Uhlmann, and Thomas Wittenberg

Automated detection of bone splinters in DEXA phantoms using deep neural networks

https://doi.org/10.1515/cdbme-2019-0071

Abstract: Dual energy radiographic imaging is a method to provide material information and can be used to differentiate between various tissue types. Dual energy X-ray absorption (DEXA) can be applied for breast density, osteoporosis or bone fracture analysis. To support radiologists with the assessment of DEXA images, machine learning can be applied. Specifically, deep convolutional neural networks (DCNNs) can be used for medical image analysis. In this work a DCNN is proposed and evaluated for automated detection of bone splinters in DEXA phantom images. The image data consists of 47 phantoms with (35) and without (12) bone splinters. Material decomposition and energy weighting results in additional image channels. Various DCNN architectures and parameters were explored. A classification rate in regions with 90 % and without 99 % bone splinters was achieved.

Keywords: DEXA, CAD, DCNN, dual energy

1 Introduction

Dual energy radiographic imaging has been proven to be a valuable method to provide spectral information of materials. It can be used to differentiate between various soft and hard tissue types in the human body. Typical applications for the use of dual energy CT (DECT) are e.g. abdominal imaging [1], differentiation of kidney stones [2], or the analysis of lung perfusion [3], whereas dual energy X-ray absorption (DEXA) can be applied for the assessment of osteoporosis [4], measurement of breast density [5], or bone analysis of fractures [6]. In order to support radiologists and physicians with the assessment of DECT or DEXA images, various machine learning approaches have been proposed and evaluated in the past, such as [7]. Recently novel methods such as deep convolutional neural networks (DCNNs) have been available for large-scale medical image analysis which can be applied for the automated assessment of dual energy imagery [8]. In order to evaluate DCNN approaches for the automated detection of bone splinters after fractions in DEXA phantom images, two different deep neural network architectures are explored and compared with each other.

***Corresponding author: Steffen Rüger,** Fraunhofer IIS Development Center X-ray Technology EZRT, Flugplatzstr. 75, 90768 Fürth, Germany, E-mail: steffen.rueger@iis.fraunhofer.de
Markus Firsching, Julija Lucic, Alexander Ennen, Norman Uhlmann, Fraunhofer IIS Development Center X-ray Technology EZRT, Flugplatzstr. 75, 90768 Fürth, Germany
Thomas Wittenberg, Fraunhofer Institute for Integrated Circuits IIS & University Erlangen, Am Wolfsmantel 33, 91058 Erlangen, Germany

2 Methodology

2.1 Phantoms

The primal data collection consist of 47 phantoms. 35 of the phantoms include bone splinters (hard tissue $\mathcal{H}$) embedded in soft tissue $\mathcal{S}$ with varying density, while 12 phantoms are solely composed of soft tissue. Quantity and type of the bone splinters vary within the phantoms. Bone splinters differ in shape, size and thickness. In total 76 bone fragments exist in the data set. A DEXA image was obtained for each phantom using a dual energy device (COMET X-ray tube with 75 kV; Hamamatsu line detector with 0.4 mm pixel pitch). The generated images consist of two channels distinguished by different energies and generated by different pre-filters on the detector side (see Fig. 1). The two energy channels are denoted as low (LE) and high (HE) energy images. Figure 2 gives examples of DEXA images with bone splinters of the LE channel.

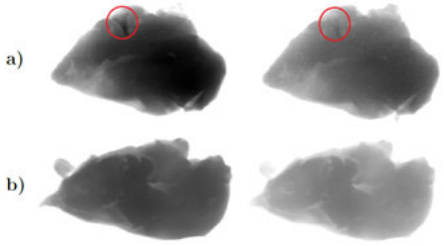

Fig. 1: Illustration of two DEXA phantom images. (a) phantom with hard (marked red) and soft tissue. (b) phantom exclusively consisting of soft tissue. Displayed in an inverted grey scale.

2.2 Data Extension

To increase the small image data collection obtained from the 47 phantoms, two approaches are applied, spectral enhancement and data augmentation.

Spectral enhancement relates to obtaining two images via dual energy absorption imaging (DEXA) instead of one mono-energetic image to increase the prior knowledge about the composition of the underlying phantom. Additionally, two further data channels are extracted by material characterization, denoted as BMD image [9] and by a weighted im-

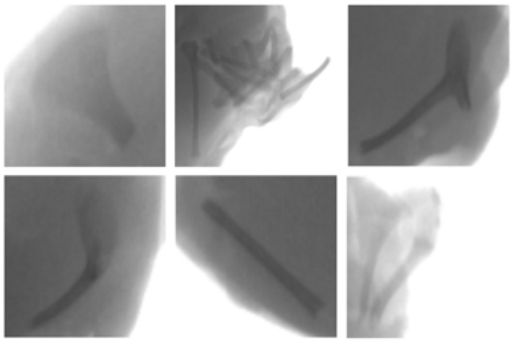

Fig. 2: Intensity-inverted bone splinter samples from DEXA images.

age (W) computed from the low and high energy images $\mathrm{W} = \mathrm{HE} - \alpha\mathrm{LE}$. The factor α was chosen so that the contrast-to-noise ratio between the hard and soft tissue regions are maximum for all phantoms.

Data augmentation methods are used as follows: (1) Mirroring the images along the horizontal axis. (2) Rotation of the images in 45 degree steps. (3) Extraction of two-dimensional pixel regions (patches) from the phantom DEXA images, denoted as the primary data augmentation method. The patches differ in the number of pixels. Patches with 15^2, 30^2 and 60^2 pixels were selected. (4) Adaption of the step size in the horizontal and vertical axis used to iteratively extract the patches from the phantom images. Step sizes correspond to one third of the respective patch size.

2.3 Data Sets

The DEXA image data collection is subdivided into three sets: training, validation and test set. This split ensures independent validation and evaluation of the deep convolutional neural network (DCNN) model based on the training process. After applying the above described data extension, the number of soft tissue samples is much higher than the number of hard tissue samples (see Tab. 1). To prevent more imbalance inside the data sets, the soft tissue phantoms are randomly divided $50 : 50$ to validation and test set. Phantoms composed of soft and hard tissue are randomly divided $60 : 20 : 20$ in training, validation and test set, with the constraint, that at least one of each hard tissue types is present in each set.

2.4 DCNN Architectures

Two different network architectures have been evaluated for the automated detection of bone splinters in DEXA phantoms (see Fig. 3 and 4). The description of each layer and the parameter terms are as follows. Parameters m and n correspond to the spatial dimension of the input patches, and varies between 15^2, 30^2 and 60^2 pixels. The variable c is a placeholder for the selected channel dimension, e.g. $c = 2$ defines using dual energy images (LE and HE) as input. The layers of the architectures can be divided in feature-extraction and decision-making parts. The feature-extraction consists of convolution ($\mathcal{C}$) and max pooling layers ($\mathcal{P}$) to sub-sample the spatial dimension of the feature space and reduce the sensitivity of the output to distortions and shifts [10]. The decision-making part incorporates one or two fully-connected layers ($\mathcal{F}$) for a binary decision making (bone splinter present or absent) as output.

Architecture $\mathcal{A}$ contains six layers and is inspired by the LeNet-5 [10]. It consists of three convolutional, two max pooling and one fully-connected layer, cf. Fig. 3. The architecture is designed in such a way that a spatial dimension of one pixel in layer $\mathcal{C}5$ is achieved. Therefore, the kernel size of the convolution layers are adapted between 3^2, 5^2 and 9^2 with a stride of 1 and a padding operation, dependent on the spatial input size (15^2, 30^2 and 60^2). The kernel used in the max pooling operation is set to 2^2 and has a stride of 2.

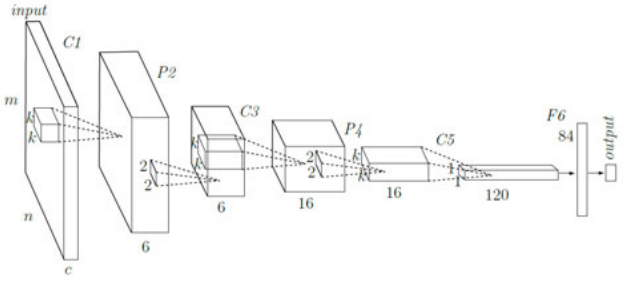

Fig. 3: Topology of architecture $\mathcal{A}$, see text for explanation.

Architecture $\mathcal{B}$ contains seven layers. The feature extracting part has three convolution layers, and two max pooling operations. The last layers are fully-connected for a binary decision making as output. Kernel size of the convolution layers are 3^2 with a stride of 1. The max pooling operation has a kernel size 2^2 with a stride of 2.

Tab. 1: The number of available phantoms and the resulting amount of samples after data extension.

Data sets:		Training			Validation			Test		
		$\mathcal{H}$	$\mathcal{S}$	Total	$\mathcal{H}$	$\mathcal{S}$	Total	$\mathcal{H}$	$\mathcal{S}$	Total
Phantoms:		23	0	23	6	6	12	6	6	12
Primary aug.:	patch size 15^2	882	16,760	17,642	186	8,656	8,842	318	8,464	8,782
	patch size 30^2	342	4,303	4,645	75	2,251	2,326	129	2,180	2,309
	patch size 60^2	170	1,084	1,254	43	596	639	61	572	633
Overall aug.:	patch size 15^2	128,342	2,416,724	2,545,066	–	–	–	–	–	–
	patch size 30^2	51,380	620,390	671,770	–	–	–	–	–	–
	patch size 60^2	23,325	161,767	185,092	–	–	–	–	–	–

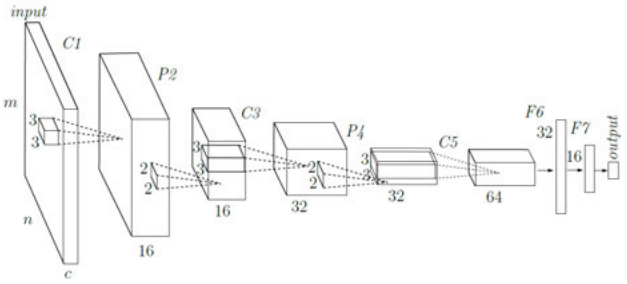

Fig. 4: Topology of architecture $\mathcal{B}$, see text for explanation.

Differences between the architectures are the available feature space in each convolution kernel and the number of layers and neurons used in the decision-making part. A continuous doubling of the feature channels in space of the convolution kernels in the architecture $\mathcal{B}$ is evaluated against the feature space of the architecture $\mathcal{A}$. In order to investigate the effect of the different number of layers and neurons in the decision part, the performance of architecture $\mathcal{A}$ is compared to architecture $\mathcal{B}$.

2.5 Training Process and Fine Tuning

For the training of the network architectures the corresponding parameters (weights and bias) have to be initialized. Initialization is performed by randomly uniform distributed data with a variance with respect to the number of input and output neurons in a certain layer [11]. Bias is set to a constant value of 0.1 to ensure that all ReLU activated neurons output in the beginning [12]. Furthermore, the set of hyperparameter is declared as epochs 5,000, batch size 128, learning rate 0.001, activation function ReLU and optimization method Adam [13–15].

Beyond data augmentation, some methods are set up to tackle overfitting the training data. (1) Normalizing the individual data sets by the Z-transform (subtraction of the mean value and division by the standard deviation calculated on the basis of the training set). (2) Using dropout to prevent the DCNN from reacting too sensitively to certain features of the input data by eliminating neurons of a layer with a chosen probability. Neuron dropout causes a different connection between the neurons in each individual optimization step. Consequently, the DCNN is not allowed to co-adapt too much between single neurons of different layers during training [16]. (3) Early stopping by freezing and saving the parameters of the architecture based on the performance of the validation set. Afterwards, the parameters are restored to evaluate the performance of the test set. (4) Applying threshold adaption of the binary decision-making. In a second instance, the threshold of the binary decision for bone splinters being present or absent is adapted. The default binary decision-making using a majority vote of the prediction probability (certainty more than 50 %) is customized, i. e. the certainty of the output that this patch belongs to one class is lowered or in most cases increased and vice versa for the remaining class. This threshold adaption is based on the performance of the validation set.

3 Results and Discussion

Sensitivity (true-positive rate, TPR), specificity (true-negative-rate, TNR) and accuracy (ACC) are used to assess performance of the DCNN-models. TPR describes the percentage of bone splinters correctly identified whereas the TNR denotes the percentage of correctly labeled soft tissue. The results are evaluated in a two level manner, on the patch level (displayed in Tab. 2) and on phantom level (shown in Tab. 3) for which the DCNN patch level output is merged back to phantom shape. Hence, the entire phantom is labeled as „bone present“ if hard tissue regions are existent or „bone absent“ if the phantom consists only of soft tissue.

Different hyperparameter spaces are declared and evaluated. The data set with primary data augmentation method (see Tab. 1) is used. Hence, a DCNN-model space Ω is determined which is taken for the training with the overall augmented data set.

$$\Omega = \left\{ \begin{array}{llll} \text{Learning rate} & 0.001, & \text{Batch size} & 128, \\ \text{Activiation func.} & \text{ReLU}, & \text{Dropout} & 50\,\%, \\ \text{Weights init.} & \text{Xavier}, & \text{Bias init.} & 0.1, \\ \text{Normalization} & \text{Z-transform}, & \text{Architecture} & \mathcal{B} \end{array} \right\}$$

For the demonstration of the results one example and its classification map is illustrated in Fig. 5. Therefore, the model training is stopped and saved on the epoch which ensures the highest accuracy based on the validation set on phantom level. Further, a threshold adaption of the binary decision-making is executed. Afterwards, the learned features of a DCNN-model are reloaded and evaluated with the test data set.

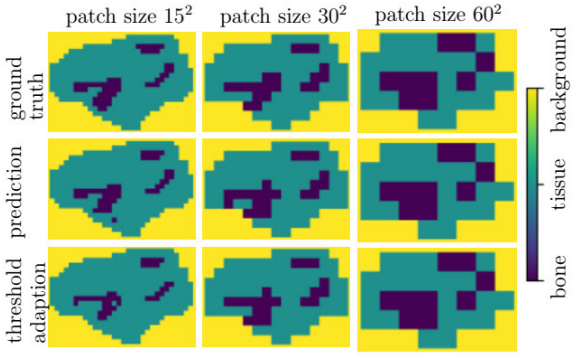

Fig. 5: Classification map of one phantom (input LE+HE images) with varying patch sizes. Patches belonging to hard tissue labels are displayed blue, soft tissue labels in green. Upper row: ground truth, center and bottom row: prediction and adaption.

A performance gain using dual energy images or their combinations as input compared to single energy images (LE or HE) as input is observed (displayed in Tab. 2 and Tab. 3). Single energy channel as input leads to the lowest results on patch and phantom level.

Tab. 2: Test results on patch level (DCNN-model space Ω and overall augmented data set for the training is used).

		LE	HE	W	BMD	LE+HE	LE+HE +BMD	BMD+W
Patch size 15^2:	TPR	61 %	35 %	43 %	55 %	75 %	72 %	69 %
	TNR	96 %	98 %	99 %	100 %	100 %	100 %	100 %
Patch size 30^2:	TPR	69 %	0 %	56 %	73 %	87 %	84 %	83 %
	TNR	98 %	100 %	99 %	99 %	99 %	99 %	99 %
Patch size 60^2:	TPR	0 %	0 %	54 %	93 %	90 %	87 %	90 %
	TNR	100 %	100 %	99 %	98 %	99 %	99 %	99 %

Tab. 3: Test results on phantom level (DCNN-model space Ω and overall augmented data set for the training is used).

		LE	HE	W	BMD	LE+HE	LE+HE +BMD	BMD+W
Patch size 15^2:	TPR	100 %	100 %	100 %	100 %	100 %	100 %	100 %
	TNR	0 %	0 %	8 %	42 %	50 %	42 %	42 %
	ACC	50 %	50 %	54 %	71 %	75 %	71 %	71 %
	Thresh.	99 %	90 %	99 %	99 %	95 %	99 %	99 %
Patch size 30^2:	TPR	92 %	83 %	100 %	92 %	100 %	100 %	100 %
	TNR	50 %	50 %	92 %	92 %	75 %	100 %	83 %
	ACC	71 %	67 %	96 %	92 %	88 %	100 %	92 %
	Thresh.	100 %	0 %	100 %	100 %	100 %	100 %	100 %
Patch size 60^2:	TPR	33 %	100 %	75 %	75 %	92 %	67 %	67 %
	TNR	67 %	50 %	88 %	88 %	96 %	83 %	83 %
	ACC	99 %	99 %	99 %	99 %	96 %	99 %	99 %
	Thresh.	83 %	0 %	100 %	100 %	100 %	100 %	100 %
	TPR	92 %	100 %	100 %	92 %	100 %	100 %	83 %
	TNR	88 %	50 %	100 %	96 %	100 %	100 %	92 %
	ACC	0 %	0 %	100 %	100 %	100 %	100 %	100 %
	Thresh.	100 %	100 %	100 %	75 %	100 %	92 %	92 %
	TPR	50 %	50 %	100 %	88 %	100 %	96 %	96 %
	TNR	99 %	99 %	20 %	71 %	84 %	88 %	93 %
	ACC	0 %	0 %	100 %	100 %	92 %	100 %	92 %
	Thresh.	100 %	100 %	75 %	83 %	100 %	92 %	100 %
	TPR	50 %	50 %	88 %	92 %	96 %	96 %	96 %
	TNR	99 %	99 %	99 %	99 %	99 %	99 %	99 %
	ACC	99 %	90 %	99 %	99 %	95 %	99 %	99 %

A slight performance gain is achieved with the weighted images as input and increased with the usage of the BMD images as input. Observing their performance on the phantom level shows that the DCNN-model has difficulty in correctly classifying some phantoms. A detailed look on the outcomes of the DCNN-model shows that the model has difficulties to classify patches correctly in which only a small region of hard tissue is present. This is also seen in the results of the TPR on the patch level. An improvement of the TPR on patch level for all evaluated input channels is achieved by enlarging the patch size. The best performance is reached for taking two channels and a patch size of 60^2 as input, the TPR is increased to 90 %. This leads to a correct classification from 24 of 24 phantoms in the test set.

4 Conclusion

Methods for automated detection of bone splinters in DEXA phantoms using deep convolutional neural networks were developed. A supervised learning approach for the classification was chosen. The data set consists of DEXA images of phantoms composed of soft and hard tissue regions or solely consisting of soft tissue. The challenge was to overcome the small number of phantoms in a way that robust splinter detection by the DCNN could be achieved. Hence, spectral enhancement and data augmentation was applied. From this the DCNN was trained for a binary classification to detect the presence or absence of bone splinters. Different architectures were designed, fine tuned and compared. Furthermore, different patch sizes and energy channels were evaluated among each other. With the usage of dual energy information in the input of the DCNN-model a TPR of 90 % and TNR of 99 % were reached on the patch level. On the phantom level, all bone splinters could be detected.

References

[1] Graser A, et al. Dual energy CT: preliminary observations and potential clinical applications in the abdomen. European Radiology 2009;19:13-23.

[2] Primak A, et al. Noninvasive differentiation of uric acid versus non–uric acid kidney stones using dual-energy CT. Acad. Radiology 2007;14:1441-1447.

[3] Thieme S, et al. Dual energy CT for the assessment of lung perfusion—Correlation to scintigraphy. EuropRad 2008;68:369-374.

[4] Blake G, Fogelman I. Role of dual-energy X-ray absorptiometry in the diagnosis and treatment of osteoporosis. J. Clinic. Densitometry 2007;10:102-110.

[5] Shepherd J, et al. Measurement of breast density with dual X-ray absorptiometry: Feasibility. Rad 2002;223:554-557.

[6] Bazzocchi A, et al. Vertebral fracture assessment by new dual-energy X-ray absorptiometry. Bone 2012;50:836-841.

[7] Li C, et al. Computer-aided diagnosis for preoperative invasion depth of gastric cancer with dual-energy spectral CT imaging. AcadRad 2015;22:149-157.

[8] Tachibana R, et al. Deep learning for electronic cleansing in dual-energy CT colonography. SPIE MediImaging 2016;9785:97851M1-7.

[9] Firsching M, et al. Multi-energy X-ray imaging as a quantitative method for materials characterization. Adv. Mater. 2011;23:2655-2656.

[10] LeCun Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE 1998;86:2278-2324.

[11] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. AISTATS 2010;9:249-256.

[12] Karpathy A. CS231n Convolutional neural networks for visual recognition. Stanford University: CS231n; 2016.

[13] Kingma D, Ba J. Adam: A method for stochastic optimization. ICLR; 2015.

[14] Glorot X, et al. Deep sparse rectifier neural networks. AISTATS 2011;15:315-323.

[15] Nair V, Hinton G. Rectified linear units improve restricted boltzmann machines. ICML; 2010.

[16] Srivastava N, et al. Dropout: a simple way to prevent neural networks from overfitting. J of Machine Learning Research 2014;15:1929-1958.