



# Automatic bone segmentation in whole-body CT images

André Klein<sup>1,2</sup> · Jan Warszawski<sup>2</sup> · Jens Hillengaß<sup>3</sup> · Klaus H. Maier-Hein<sup>1</sup>

Received: 26 June 2018 / Accepted: 5 November 2018 / Published online: 13 November 2018  
© CARS 2018

## Abstract

**Purpose** Many diagnostic or treatment planning applications critically depend on the successful localization of bony structures in CT images. Manual or semiautomatic bone segmentation is tedious, however, and often not practical in clinical routine. In this paper, we present a reliable and fully automatic bone segmentation in whole-body CT scans of patients suffering from multiple myeloma.

**Methods** We address this problem by using convolutional neural networks with an architecture inspired by the U-Net [17]. In this publication, we compared three training procedures: (1) training from 2D axial slices, (2) a pseudo-3D approach including axial, sagittal and coronal slices and (3) an approach where the network is pre-trained in an unsupervised manner.

**Results** We evaluated the method on an in-house dataset of 18 whole-body CT scans consisting of 6800 axial slices, achieving a dice score of 0.95 and an intersection over union (IOU) of 0.91. Furthermore, we evaluated our method on the dataset used by Pérez-Carrasco et al. (Comput Methods Progr Biomed 156:85–95, 2018). The data and the ground truth have been made publicly available. The proposed method outperformed the other methods, obtaining a dice score of 0.92 and an IOU of 0.85.

**Conclusion** These promising results could facilitate the evaluation of bone density and the localization of focal lesions in the future, with a potential impact on both disease staging and treatment planning.

**Keywords** Bone segmentation · U-Net · Deep learning · Computed tomography · Multiple myeloma

## Introduction

High-quality and fast automatic bone segmentation in CT images is important for analysis, staging and treatment planning of many diseases like multiple myeloma. The various, irregular shapes of bones and their homogeneous internal structure make automatic bone segmentation a significant challenge, despite several years of research already invested into the topic [9]. This is aggravated by the fact that CT scans in clinical routine are often captured with a low radiation dose which leads to inferior image quality.

In general, one differentiates between four categories of bone based on its shape: long bones, short bones, flat bones

and irregular bones [4]. Regardless of the category, various tissue types are referred to as bone. We differentiate between cortical (compact) bone, cancellous (trabecular, spongy) bone and bone marrow, as shown in Fig. 1. The density of bone tissue is measured in Hounsfield units (HUs), which can vary significantly between the tissue types described above. Cortical bone is the most dense and solid part of the bone, represented by high HU. Cancellous bone and bone marrow are less dense, with HU being more similar to those of soft tissue like muscle. Pathological changes in the bone, as they occur in patients suffering from multiple myeloma, can influence the density and therefore the HU of bone tissue [6].

The gold standard, slice-by-slice hand contouring, is very time-consuming, tedious and error-prone [2]. Reliable fully automatic bone segmentation has therefore been of great interest to research for a long time. Despite the large number of different approaches and studies regarding the topic, bone segmentation is still considered an ongoing problem in several aspects [9].

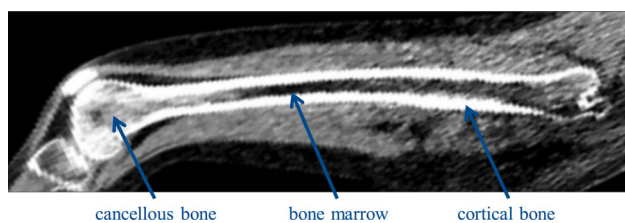
In some approaches, it is considered as a local problem, concentrating on specific bones or body regions. Long bones like the femur, for example, are focused on by var-

✉ André Klein  
andre.klein@dkfz.de

<sup>1</sup> Division of Medical Image Computing, Deutsches Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 581, Heidelberg, Germany

<sup>2</sup> Medical Faculty, University of Heidelberg, Heidelberg, Germany

<sup>3</sup> Department of Medicine, Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA



**Fig. 1** CT scan of the femur. The cortical bone appears white and surrounds the less dense cancellous bone and the bone marrow

ious authors, such as Krčah et al. [9] and Younes et al. [9]. Younes et al. proposed primitive shape recognition and statistical shape models. Pinheiro et al. took a more general approach by not focusing on a specific bone but on a user-defined region of interest. They apply a level-set-based protocol [16]. In addition, there are also approaches that deal with whole-body scans. Pérez-Carrasco et al. applied continuous max-flow optimization [15] as well as histogram-based energy minimization [14]. Further approaches are based on region growing, intensity thresholding [2], energy minimizing spline curves, edge detection or combinations of these algorithms. Sometimes, expensive pre- and post-processing steps are necessary or the algorithms are dependent on a specific initialization [12].

Deep learning algorithms have become the methodology of choice in many areas of automatic medical image segmentation issues [10]. Yet, their performance on bone segmentation tasks remains to be evaluated. Some initial work can be found in the “Bone Segmenter” project by Kevin Mader.<sup>1</sup>

In this paper, we present our most recent efforts on bone segmentation on whole-body CT images. We propose a network based on the U-Net architecture by Ronneberger et al. [17] as it has become a commonly used benchmark in medical image segmentation. We compare three different training strategies with the goal to locate and segment cortical and cancellous tissue as well as the bone marrow in whole-body scans of patients with multiple myeloma, regardless of their shape.

## Materials and methods

### Data

In this paper, we rely on an in-house dataset as well as a publicly available one. We train and evaluate our method on both datasets independently.

**In-house dataset** The in-house dataset consists of 53 low-quality low-dose whole-body CT scans that were captured

<sup>1</sup> <https://github.com/4Quant/Bone-Segmenter>.

as part of a PET/CT study during standard assessment for patients with multiple myeloma. The acquisition device was a Biograph 128 PET/CT Scanner by Siemens. The spacing is equal for all scans ( $0.98 \times 0.98 \times 4 \text{ mm}^3$ ). Axial slices are  $512 \times 512$  pixels, while each scan has between 380 and 450 slices. For 18 of these scans, a ground truth segmentation was performed by a medical expert using the segmentation plug-in of the Medical Imaging Interaction Toolkit (MITK) [13]. The expert adapted a base segmentation created for each scan by an intensity threshold. We perform a sixfold cross-validation using 12 scans for training ( $\approx 4800$  axial slices), 3 for validation ( $\approx 1200$  axial slices) and 3 for testing ( $\approx 1200$  axial slices). The unlabeled scans are used for pre-training as described in “Training” section.

**Publicly available dataset** For better comparison of our method, we trained and validated our network on 2D axial slices from the publicly available dataset by Pérez-Carrasco et al. [14]. The dataset is comprised of 27 CT volumes from 20 patients. It was acquired by a Helical CT by Philips Medical Systems, with a slice size of  $512 \times 512$  and a spacing of ( $0.78 \times 0.78 \times 5 \text{ mm}^3$ ). Fifteen CT volumes were used for training, 3 for validation and 9 for testing.

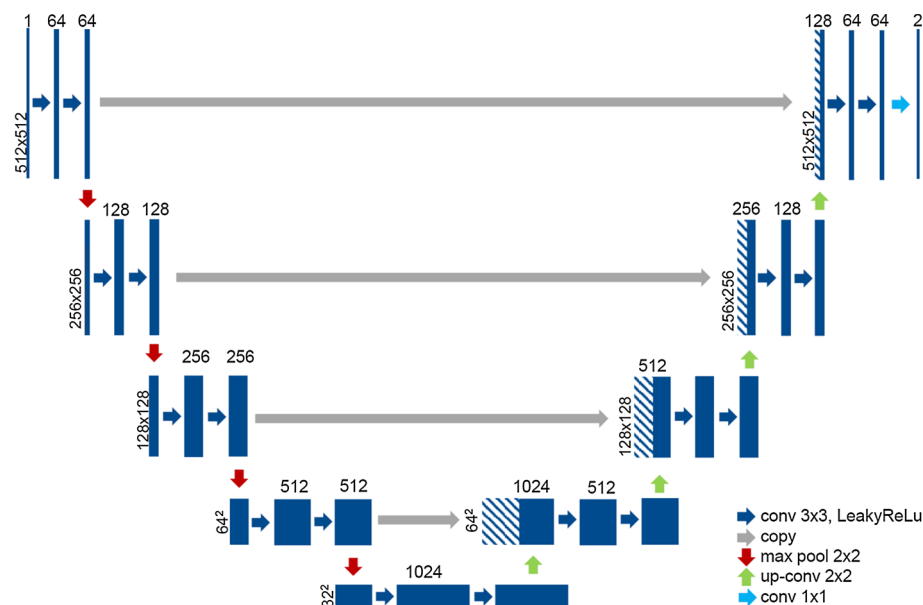
### Architecture

The convolutional neural network we use is an adapted version of the U-Net that was initially proposed by Ronneberger et al. [17]. The architecture follows the encoding–decoding principle. It comprises an analysis pathway for context aggregation of increasingly abstract input representations and a synthesis pathway that combines the semantic information from deeper layers with spatial information from shallower layers. Our model is shown in Fig. 2. We designed the network to process 2D input images of  $512 \times 512$  pixels. Our model uses padded convolutions with a kernel size of 3 and input stride 2 to keep the spatial output dimensions equal to the input. We replace the rectified linear units (ReLU) with leaky ReLU nonlinearities [11] with a negative slope of  $10^{-2}$  [8]. This selection is based on empirical experiences and needs to be validated in the future.

### Training

In this chapter, we present three different training strategies: (1) training from 2D axial slices, (2) a pseudo-3D approach including axial, as well as sagittal and coronal slices and (3) an approach, where the network is pre-trained in an unsupervised manner. All of them use data augmentation to prevent our network from overfitting and to more efficiently train given the amount of training data [17]. We make use of  $\pm 10^\circ$  rotations around the axial axis, as well as randomly vertical mirroring and scaling from 60 to 140%. Furthermore,

**Fig. 2** The architecture is based on the U-Net architecture as proposed by Ronneberger et al. [17]. It consists of an analysis path that captures semantic information and a symmetric synthesis path that enables precise localization information [17]



we apply random elastic deformations. After consulting a medical expert, we chose the augmentation parameters to be as aggressive as possible [8], while still ensuring realistic augmented medical images. All the augmentation techniques are applied on-the-fly with our own in-house framework which is publicly available at <https://github.com/MIC-DKFZ/batchgenerators>. The network is trained with a batch size of 8 and input patches of  $512 \times 512$ . In the field of bone segmentation, the amount of foreground and background pixels is imbalanced. We approach this issue by using a combination of cross-entropy and dice loss:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{dice}} \tag{1}$$

The dice loss is implemented as follows [7]:

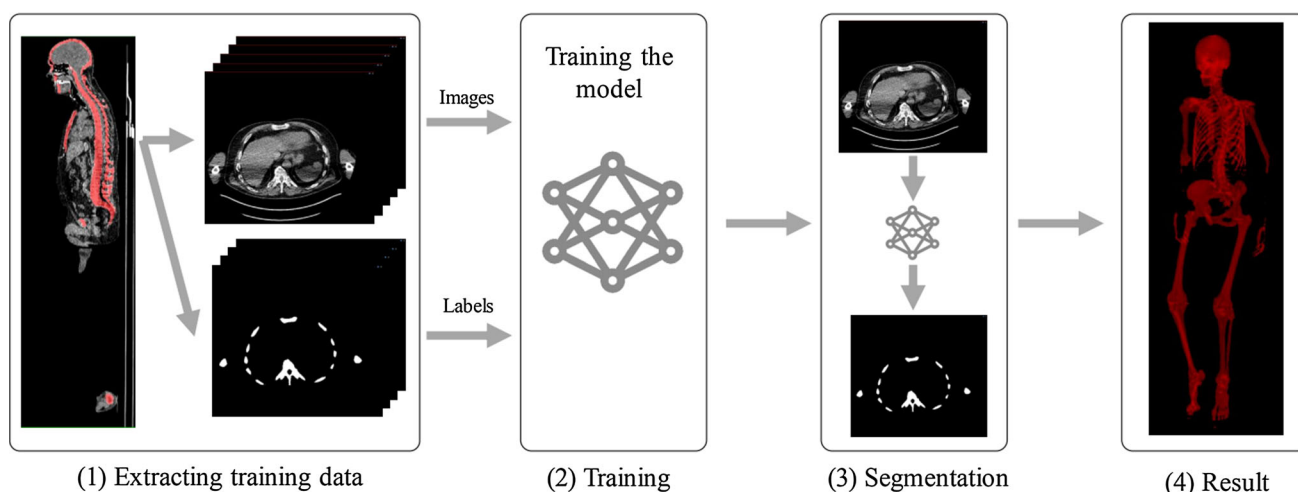
$$\mathcal{L}_{\text{dice}} = - \frac{2}{|K|} \sum_{k \in K} \frac{\sum_i u_{i,k} v_{i,k}}{\sum_i u_{i,k} + \sum_i v_{i,k}} \tag{2}$$

where  $u$  is the softmax output of the network and  $v$  is a one-hot encoding of the ground truth segmentation map [7].  $u$  and  $v$  are the same size ( $i \times k$ ), with  $i$  being the number of pixels in the training patch and  $k \in K$  being the number of classes. We use an ADAM optimizer with a learning rate of  $10^{-4}$ ,  $\beta_1 = 0.5$  and  $\beta_2 = 0.999$  for our training. Our network is trained for 100 epochs.

**2D training from axial slices** We use axial slices of  $512 \times 512$  pixels as input (see Fig. 3). The slices are extracted randomly from whole-body CT scans.

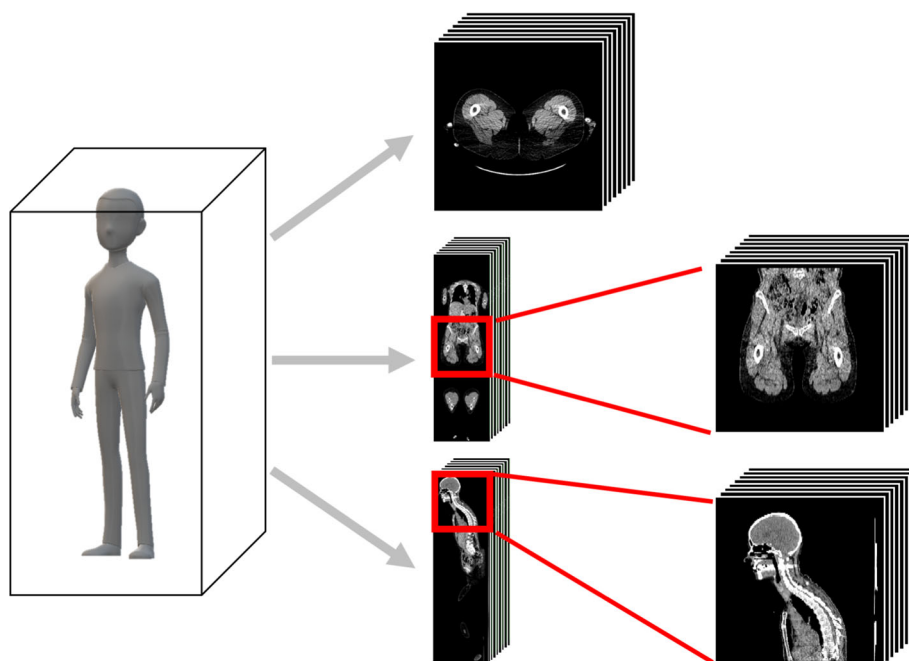
**Pseudo 3D training** The standard training does not take into account any 3D information. To achieve that, we train our model in what we call a “pseudo-3D” way. We alternate the input batches between batches of axial, sagittal and coronal slices. This way the network learns to segment bony structures independent of the viewing direction. This training procedure is inspired by the work of Wasserthal et al. [18]. In contrast to their work, our 3D datasets are not equally spaced. We resample the input images to a  $1 \times 1 \times 1 \text{ mm}^3$  spacing. In dealing with whole-body scans, the datasets are usually  $> 512$  pixels in cranial–caudal direction. We extract input images of  $512 \times 512$  pixels from the whole sagittal and coronal slices. During training, the position of the extracted patch is chosen randomly. For testing, we use a  $512 \times 512$  sliding window to segment the whole image (see Fig. 4). A separate segmentation is done for each viewing direction. The three outputs per volume are merged using the mean to generate the final segmentation.

**Unsupervised pre-training** As shown by Erhan et al., unsupervised pre-training has a regularization effect and adds robustness to deep architectures [5], thus reducing model variance [19]. We pre-train our model as an auto-encoder by disabling the skip connections, as they could potentially act as shortcuts, leading to a decreased pre-training effect on deeper levels. Furthermore, we replaced the output softmax layer with an output reconstruction layer that performs a  $1 \times 1$  convolution. The goal of this end-to-end pre-training is for the network to learn to reconstruct the received input image. For the pre-training, we use a mean-squared-error loss as proposed by Wiehmann et al. [19]. We pre-train our network for 30000 iterations.



**Fig. 3** Our pipeline: (1) Extracting batches of  $512 \times 512$  images and corresponding ground truth from 3D CT volumes. (2) Training the model. (3) Segmentation of  $512 \times 512$  images with the trained model. (4) Generation of 3D whole-body bone segmentation from 2D slices

**Fig. 4** For the pseudo-3D training, we extract either axial, sagittal or coronal batches from our whole-body CT scans. Because the scans were resampled to a  $1 \times 1 \times 1 \text{ mm}^3$ , we crop patches of size  $512 \times 512$  from the slices to fit our network architecture



## Segmentation

To test our method, we segment sequential batches extracted from a test CT scan and concatenate them to a whole-body CT bone segmentation. The result is then compared to the ground truth. For the pseudo-3D approach, the test scans have to be segmented three times: in axial, sagittal and coronal orientation. Doing this, we end up with three different segmentations for the dataset. To generate the final whole-body segmentation, we calculate the mean likelihood for each pixel from the values of the output softmax layer.

For the comparison of our method on the publicly available dataset, we only trained and segmented on 2D axial slices,

as the difference between the proposed training procedures was minimal.

## Results

We compare the segmentation of our networks to the ground truth by calculating the following metrics: dice score, IOU (Jaccard index), sensitivity, specificity, positive predictive value (PPV) and accuracy. Each of these metrics highlights a different aspect of the quality of the segmentation [3].

**In-house dataset** For our in-house dataset, we further compare the results to a naïve approach of thresholding and a

set of morphological operations. The results of the proposed segmentation algorithms are shown in Fig. 5. Because the in-house dataset is rather small, we chose to calculate statistics across all CV folds, i.e., test set predictions of each fold were collected and fused to a final test set containing all patients. The confidence bounds show the inter-patient standard deviation. The network, trained from 2D axial slices, performed best, achieving a dice score of  $0.95 \pm 0.01$  and an intersection over union (IOU) of  $0.91 \pm 0.02$ . Pre-training our network with unlabeled whole-body CT scans did neither improve nor worsen the results. It can be observed, though, that the pre-training helps the network to converge after fewer iterations. The validation loss with pre-training converges at a plateau of about 0.96 after about 40 epochs. Without pre-training, the validation loss converges at the same value after about 80 epochs. The pseudo-3D network achieved a dice score of  $0.92 \pm 0.01$  and an IOU of  $0.85 \pm 0.02$ . In comparison, the naïve approach achieved a dice score of  $0.70 \pm 0.05$  and an IOU of  $0.54 \pm 0.07$ .

To emphasize the effect of the pre-training, we trained the 2D networks with a lower amount of training data ( $n = 3$  scans and  $n = 6$  scans). The gain in the results on our test set is minimal when increasing the amount of training data.

The proposed networks, as well as the naïve approach, work well for cortical bone due to its high HU values. Examples for the naïve approach are shown in the second column of Fig. 6. The main issues arise when segmenting spongy bone and bone marrow. As these tissues are less dense, their HU values are more similar to those of soft tissue. As expected, a thresholding-based approach does not segment these tissues well. It also often mistakes the table in the images as bony structures (see Fig. 6). The results of the segmentation strongly depend on the chosen threshold. The mentioned issues are partly solved by our learning-based approach as it does not rely on HU values alone. However, the performance on more complex body regions like the skull and the chest is still challenging (see Fig. 5).

We compared the metrics of different body regions. Our networks achieve the best results on the legs, followed by the pelvis and the upper body. The results for the head are slightly worse. This is due to very thin bone structures in the skull and artifacts caused by tooth crowns. Another difficult task for all approaches was patients with artificial joints, such as hips or knees, as they also cause artifacts in the CT scans.

Training time was approximately two days per network. The segmentation of a whole-body CT scan ( $512 \times 512 \times 400$ ) slices took about 50 s for the 2D axial and about 9 min and 30 s for the pseudo-3D segmentation on an NVIDIA Titan X GPU.

**Publicly available dataset** We compared our proposed method with the state-of-the-art techniques on the publicly available dataset by Pérez-Carrasco et al. [14]: thresholding,

pixel-value-based convex relaxation [1, 14], histogram-based convex relaxation, hybrid-level-set model technique [20] and histogram-based energy minimization [14].

Table 1 shows a comparison of the performance metrics evaluated for the proposed model and benchmark algorithms. As can be seen, the proposed method outperformed the other methods in five out of six metrics. The only exception was sensitivity, in which the histogram-based energy minimization achieved the highest score. We assume that the proposed method achieves a higher specificity than the histogram-based energy minimization. Due to rounding inaccuracy, this cannot be observed in Table 1. Sensitivity and specificity rely on the total amount of pixels. Because of the class imbalance in the segmentation task, a few pixels cause a rather big difference in sensitivity, but a minimal difference in specificity results. This is substantiated by the fact that the proposed method performs better in the other metrics.

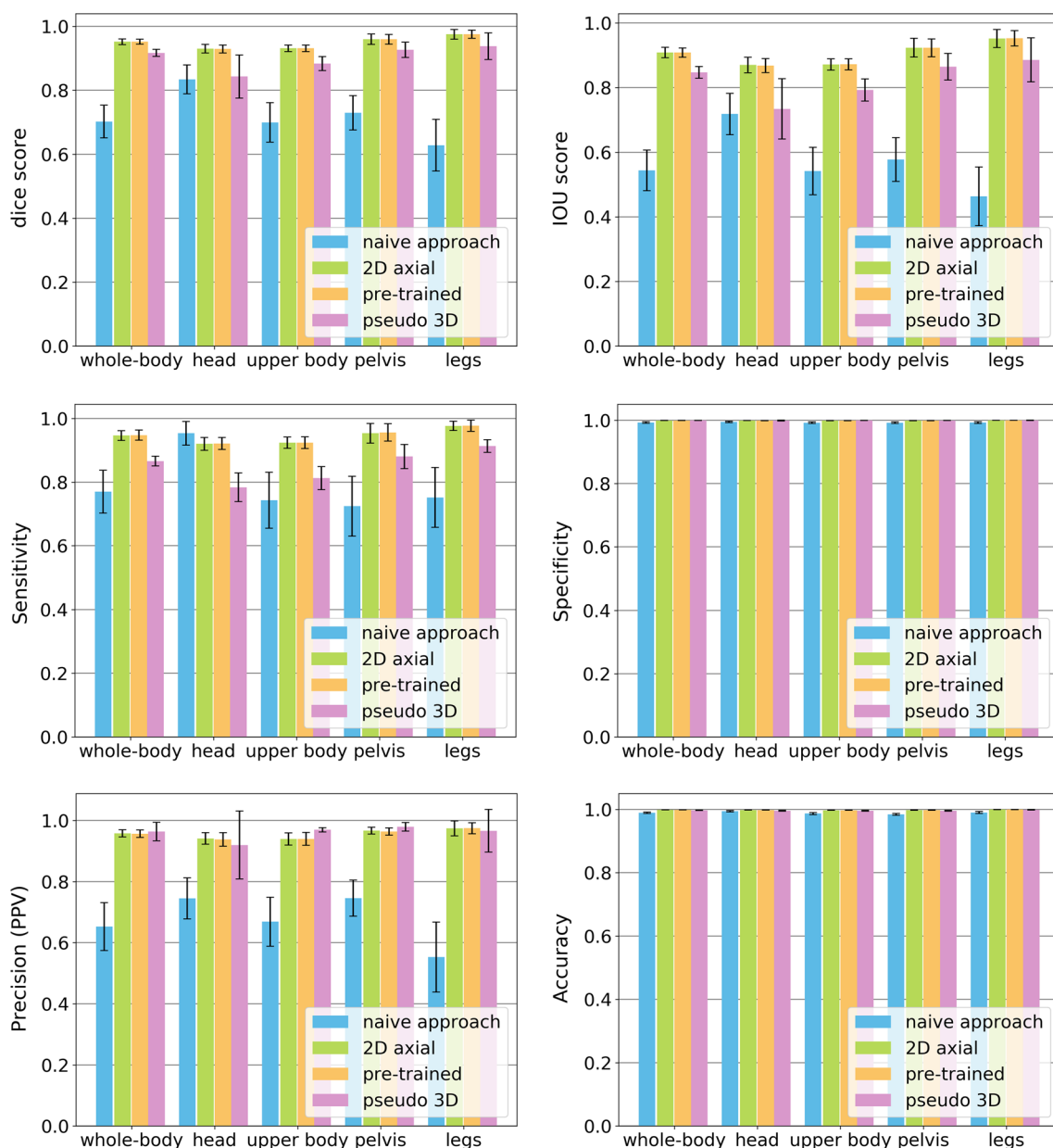
## Discussion

In this paper, we present a deep learning approach for the simultaneous segmentation of long, short, flat and irregular bones including cortical, spongy and bone marrow structures. We use a deep convolutional neural network inspired by U-Net that we train from scratch using an in-house dataset, as well as a publicly available one, extensive data augmentation, and a combination of a cross-entropy and a dice-loss formulation. Furthermore, we examine the effect of unsupervised pre-training and pseudo-3D training on the segmentation results.

We evaluated six different metrics for the proposed training procedures. However, metrics like sensitivity, specificity and accuracy do not take into account the class imbalance prevalent in the data. We achieve dice scores of 0.95 and IOU 0.91 on whole-body CT scans for our best network. We evaluated the performance on different body regions for our in-house whole-body dataset. The best results were achieved on large bones like the femur. As expected, the segmentation of smaller bones like the ribs is more challenging. This is related to the fact that only small pieces of each rib are visible on each slice and that they are surrounded by more complex tissue combinations in comparison with the femur (see the third row of Fig. 6). Additionally, the placement of the arms during the CT scan can lead to attenuation of the signal in some areas and thereby cause segmentation errors. A common practice is to arrange the arms—which is angled next to the head, to prevent such attenuation effects. Because we took CT scans of patients suffering from multiple myeloma, the positioning of the patient had to be adjusted in a way that is comfortable and does not cause pain.

The most challenging body region for our networks to segment is the skull. This is caused by various small and irregular





**Fig. 5** Comparison of dice, IOU, sensitivity, specificity, PPV and accuracy of for different approaches and different body regions

shaped bones in that area. Furthermore, most of the patients in our dataset have tooth crowns. These cause strong artifacts that make the segmentation of bone very challenging not only to our algorithm but to the medical expert as well. The ambiguous image information leads to noisy labels, because the medical expert cannot annotate the data in a well-defined way. Although these artifacts are present in many patients, they present a significant problem because evaluation of the segmentation is hard due to label noise. On slices like the one shown in the second row of Fig. 6, the upper jaw was not segmented because the artifacts make it nearly impossible to distinguish between bone and surrounding tissue. While

the naïve approach based on an intensity threshold clearly over-segments, the segmentation of our network looks like a possible segmentation of the jaw bone. Because of the noisy ground truth in these areas, such circumstances complicate the validation of our approach.

Another source of artifacts is artificial joints. Only two out of 18 patients in our in-house training dataset had these (see the fifth row of Fig. 6). Artificial joints were not segmented by our medical expert in our ground truth as they do not consist of bone tissue. The artifacts often have similar HU values as cortical bone which makes the task of bone segmentation difficult on the according slices. As these problematic slices

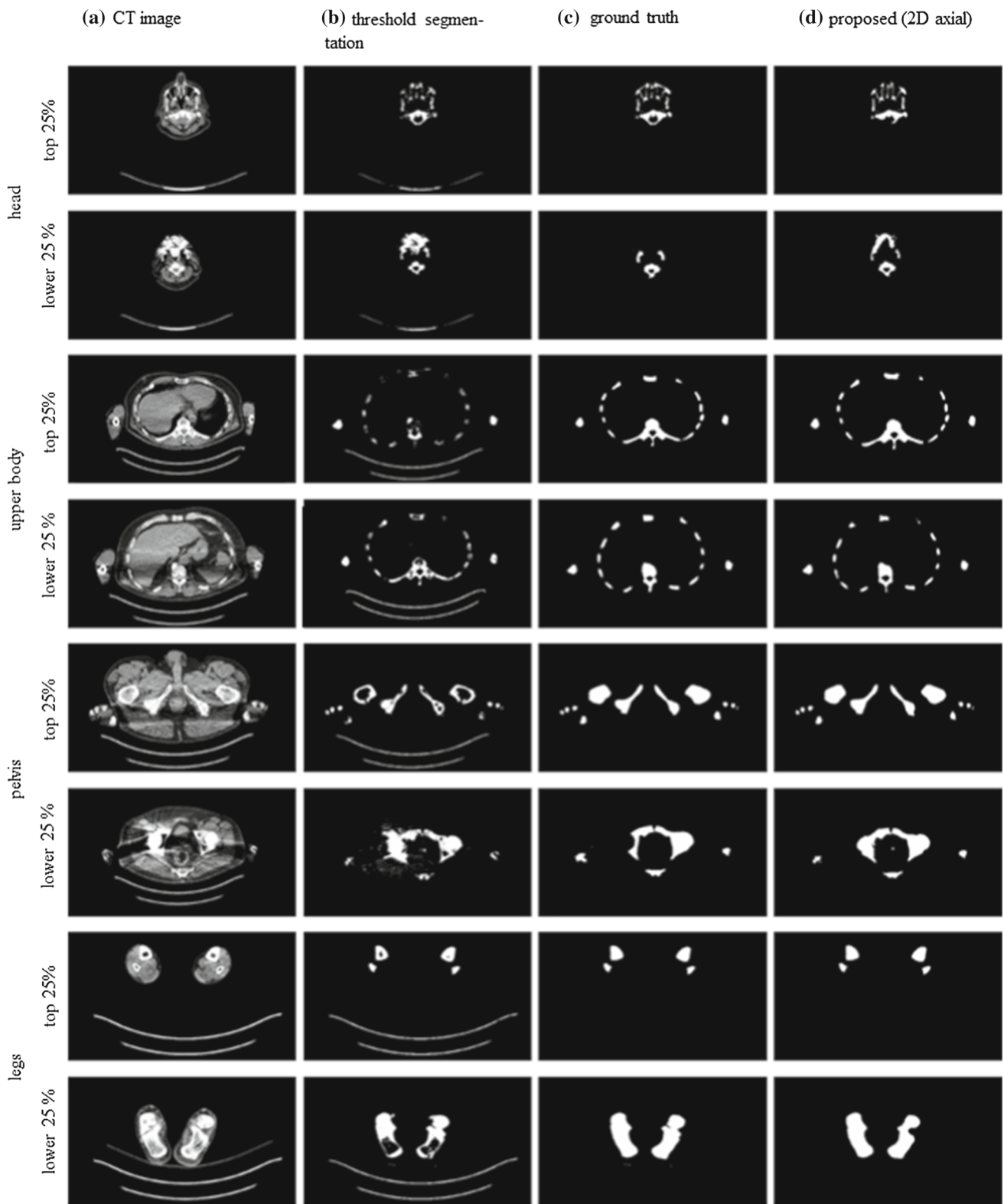


Fig. 6 Qualitative results for body regions

**Table 1** Comparison of performance metrics evaluated for the proposed model and benchmark algorithms

	Thresholding	Pixel-value-based relaxation [1]	Histogram-based convex relaxation	Hybrid-level-set model [20]	Histogram-based energy minimization [14]	Proposed
Dice	0.79 ± 0.12	0.82 ± 0.19	0.83 ± 0.19	0.73 ± 0.12	0.88 ± 0.14	<b>0.92 ± 0.05</b>
Jaccard (IoU)	0.66 ± 0.15	0.73 ± 0.23	0.73 ± 0.16	0.61 ± 0.21	0.80 ± 0.19	<b>0.85 ± 0.08</b>
Sensitivity	0.71 ± 0.14	0.78 ± 0.23	0.85 ± 0.17	0.85 ± 0.09	<b>0.94 ± 0.15</b>	0.91 ± 0.08
Specificity	1.00 ± 0.00	1.00 ± 0.00	0.99 ± 0.00	0.99 ± 0.01	1.00 ± 0.00	1.00 ± 0.00
PPV (precision)	0.91 ± 0.11	0.90 ± 0.14	0.84 ± 0.10	0.69 ± 0.25	0.90 ± 0.15	<b>0.94 ± 0.04</b>
Accuracy	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	<b>1.00 ± 0.00</b>

The proposed method outperformed the other methods in five out of six metrics. The best result for each metric is highlighted in bold

The comparison results presented in this table were taken from the paper of Pérez-Carrasco et al [14]. The methods were not re-implemented for this paper

are not represented sufficiently frequent in our dataset, it is difficult for the proposed method to learn how to adequately handle such situations.

We compared the mean metrics of each patient, performing a *t* test, to determine whether there is a significant difference between the results. The margin between the training strategies with and without pre-training is minimal (*p* value=0.99). We expected the effects of the unsupervised pre-training to be more prominent if the networks are trained with less training data. This was not the case probably due to the fact that the results of the 2D axial approach are already at a level of saturation. The pseudo-3D approach performed worst compared to the other two deep learning methods (*p* value  $\ll$  0.01). This, however, might be related to the acquisition of the ground truth: The medical expert created the ground truth segmentations on axial slices. The spacing of our in-house dataset in cranial–caudal direction is four times bigger, causing the ground truth segmentation on coronal and sagittal slices to be much noisier than the axial ones. All three training strategies perform significantly better than the naïve approach of thresholding and morphological operations (*p* value  $\ll$  0.01).

Many different bone segmentation approaches have been published so far. It is not easy to provide a fair comparison of the different algorithms, as a lot of the work is focused on restricted problems like the segmentation of specific bony structures. We compared our method to benchmark algorithms on the publicly available dataset by Pérez-Carrasco et al. The dataset is very small for a deep learning approach, but the results were still very promising.

We plan to evaluate the possibility to transfer our method to the clinic. For this, the robustness must be further evaluated. Both datasets used in this paper were acquired with a single scanner—each. To ensure the robustness of our method, we would need a more heterogeneous dataset to train on. Such a dataset could be established in the future to establish a more general bone segmentation method that applies to a variety of scanners and different levels of image

quality. Furthermore—on our in-house dataset—we trained on scans of patients suffering from multiple myeloma. As the disease decomposes the bone, this could also affect the bone segmentation. We plan to validate our approach on a wider range on image data and we will continue to further expand our reference dataset.

**Acknowledgements** This research was supported by the International Myeloma Foundation.

## Compliance with ethical standards

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

**Ethical standard** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Bae E, Yuan J, Tai XC (2011) Global minimization for continuous multiphase partitioning problems using a dual approach. *Int J Comput Vis* 92(1):112–129
- Buie HR, Campbell GM, Klinck RJ, MacNeil JA, Boyd SK (2007) Automatic segmentation of cortical and trabecular compartments based on a dual threshold technique for in vivo micro-ct bone analysis. *Bone* 41(4):505–515
- Chang HH, Zhuang AH, Valentino DJ, Chu WC (2009) Performance measure characterization for evaluating neuroimage segmentation algorithms. *Neuroimage* 47(1):122–135
- Clarke B (2008) Normal bone anatomy and physiology. *Clin J Am Soc Nephrol* 3(Supplement 3):S131–S139
- Erhan D, Bengio Y, Courville A, Manzagol PA, Vincent P, Bengio S (2010) Why does unsupervised pre-training help deep learning? *J Mach Learn Res* 11(Feb):625–660
- Hillengass J, Delorme S (2012) Multiples Myelom: Aktuelle Empfehlungen für die Bildgebung. *Der Radiologe* 52(4):360–365. <https://doi.org/10.1007/s00117-011-2257-0>



7. Isensee F, Jaeger P, Full PM, Wolf I, Engelhardt S, Maier-Hein KH (2017) Automatic cardiac disease assessment on cine-mri via time-series segmentation and domain specific features. arXiv preprint [arXiv:1707.00587](https://arxiv.org/abs/1707.00587)
8. Isensee F, Kickingereder P, Wick W, Bendszus M, Maier-Hein KH (2017) Brain tumor segmentation and radiomics survival prediction: contribution to the brats 2017 challenge. In: 2017 International MICCAI BraTS challenge
9. Krčah M, Székely G, Blanc R (2011) Fully automatic and fast segmentation of the femur bone from 3d-ct images with no shape prior. In: 2011 IEEE international symposium on biomedical imaging: from nano to macro, pp 2087–2090
10. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, van der Laak JA, van Ginneken B, Sánchez CI (2017) A survey on deep learning in medical image analysis. arXiv preprint [arXiv:1702.05747](https://arxiv.org/abs/1702.05747)
11. Maas AL, Hannun AY, Ng AY (2013) Rectifier nonlinearities improve neural network acoustic models. In: Proceedings of the ICML, vol 30, p 3
12. Malladi R, Sethian JA, Vemuri BC (1995) Shape modeling with front propagation: a level set approach. *IEEE Trans Pattern Anal Mach Intell* 17(2):158–175
13. Nolden M, Zelzer S, Seitel A, Wald D, Müller M, Franz AM, Maleike D, Fangerau M, Baumhauer M, Maier-Hein L, Maier-Hein KH, Meinzer HP, Wolf I (2013) The medical imaging interaction toolkit: challenges and advances. *Int J Comput Assist Radiol Surg* 8(4):607–620. <https://doi.org/10.1007/s11548-013-0840-8>
14. Pérez-Carrasco JA, Acha B, Suárez-Mejías C, López-Guerra JL, Serrano C (2018) Joint segmentation of bones and muscles using an intensity and histogram-based energy minimization approach. *Comput Methods Progr Biomed* 156:85–95
15. Pérez-Carrasco JA, Acha-Piñero B, Serrano C (2015) Segmentation of bone structures in 3d ct images based on continuous max-flow optimization. *Med Imaging Image Process*. <https://doi.org/10.1117/12.2082139>
16. Pinheiro M, Alves J (2015) A new level-set-based protocol for accurate bone segmentation from ct imaging. *IEEE Access* 3:1894–1906
17. Ronneberger O, Fischer P, Brox T (2015) U-net: Convolutional networks for biomedical image segmentation. In: International conference on medical image computing and computer-assisted intervention (MICCAI), pp 234–241
18. Wasserthal J, Neher PF, Maier-Hein KH (2018) Tractseg—fast and accurate white matter tract segmentation. arXiv preprint [arXiv:1805.07103](https://arxiv.org/abs/1805.07103)
19. Wiehman S, Kroon S, De Villiers H (2016) Unsupervised pre-training for fully convolutional neural networks. In: Pattern Recognition Association of South Africa and robotics and mechatronics international conference (PRASA-RobMech). IEEE, pp 1–6
20. Zhang Y, Matuszewski BJ, Shark LK, Moore CJ (2008) Medical image segmentation using new hybrid level-set method. In: Fifth international conference biomedical visualization: information visualization in medical and biomedical informatics. IEEE, pp 71–76