International Journal of Computer Assisted Radiology and Surgery (2023) 18:2091–2099
https://doi.org/10.1007/s11548-023-02957-4

ORIGINAL ARTICLE

Improved distinct bone segmentation in upper-body CT through multi-resolution networks

Eva Schnider[1] · Julia Wolleb[1] · Antal Huck[1] · Mireille Toranelli[2] · Georg Rauter[1] · Magdalena Müller-Gerbl[2] · Philippe C. Cattin[1]

Received: 10 January 2023 / Accepted: 9 May 2023 / Published online: 20 June 2023

Abstract

Purpose Automated distinct bone segmentation from CT scans is widely used in planning and navigation workflows. U-Net variants are known to provide excellent results in supervised semantic segmentation. However, in distinct bone segmentation from upper-body CTs a large field of view and a computationally taxing 3D architecture are required. This leads to low-resolution results lacking detail or localisation errors due to missing spatial context when using high-resolution inputs.

Methods We propose to solve this problem by using end-to-end trainable segmentation networks that combine several 3D U-Nets working at different resolutions. Our approach, which extends and generalizes HookNet and MRN, captures spatial information at a lower resolution and skips the encoded information to the target network, which operates on smaller high-resolution inputs. We evaluated our proposed architecture against single-resolution networks and performed an ablation study on information concatenation and the number of context networks.

Results Our proposed best network achieves a median DSC of 0.86 taken over all 125 segmented bone classes and reduces the confusion among similar-looking bones in different locations. These results outperform our previously published 3D U-Net baseline results on the task and distinct bone segmentation results reported by other groups.

Conclusion The presented multi-resolution 3D U-Nets address current shortcomings in bone segmentation from upper-body CT scans by allowing for capturing a larger field of view while avoiding the cubic growth of the input pixels and intermediate computations that quickly outgrow the computational capacities in 3D. The approach thus improves the accuracy and efficiency of distinct bone segmentation from upper-body CT.

Keywords Multi-resolution · Distinct bone segmentation · Deep learning

Introduction

Segmentation of bones is used in bone disease diagnosis, in image-based assessment of fracture risks [1], bone-density [2], for planning and navigation of interventions [3], and for post-treatment assessment.

Bone tissue segmentation from CT has been shown to work well using slice-wise 2D CNN-based segmentation algorithms [4–6]. The tasks and solutions become more varied when moving from bone tissue segmentation to distinct bone segmentation (our task) where we distinguish individual bones. Vertebrae segmentation has gained much attention, with many of the algorithms using multi-stage approaches and leveraging the sequential structure of the spine [7]. Rib segmentation has been tackled by [8], who use a point cloud approach targeted at leveraging their dataset's spatial sparsity. Carpal bone segmentation is performed from X-rays of hands that were placed on a flat surface [9].

Simultaneous segmentation of distinct bones of multiple groups is still relatively little studied. A cascade of a bone tissue segmentation and a distinct bone segmentation network have been used by [10] to segment eight upper and lower limb bones from whole-body CT. Fu et al. [11] segment 62 different bones from upper-body CT using an atlas-based approach and kinematic joint models. Lindgren Belal et al. [12] use a multi-stage approach with a localisation network, shape models, and a segmentation network to segment 49 distinct

✉ Eva Schnider
eva.schnider@unibas.ch

[1] Department of Biomedical Engineering, University of Basel, Hegenheimermattweg 167B, 4123 Allschwil, Switzerland

[2] Department of Biomedicine, Musculoskeletal Research, University of Basel, Basel, Switzerland

bones of the upper body. Segmentation of bones of different groups in one shot can be used as a starting point for more fine-grained atlas segmentations [11], or as a guide for a follow-up inner organ segmentation [13]. Segmenting multiple structures at once can also be beneficial for the segmentation accuracy [14], found their network trained on multiple bone classes to outperform the one-class networks.

The region of interest in upper-body or full-body CT scans is typically larger than the possible input sizes of 3D convolutional neural networks (CNNs). As a result, the input needs to be sampled as patches, restricting the input field of view to the patch size. This problem exacerbates with the development of CT scanners that produce ever more highly resolved images. While a higher resolution allows for capturing more fine-grained details, it covers smaller body areas within a fixed-size input patch.

In order to extend the field of view, larger input patches can be sampled. Using bigger patches, i.e. more input pixels does not increase the number of trainable parameters in a fully connected network, but it does increase the number of necessary intermediate computations. Doubling the patch size in all three dimensions leads to at least eight times more forward- and backward computations, which are taxing for the generally scarce GPU memory. Countermeasures fall into two categories. (A) keeping the resolution and input pixel size high, but reducing the computational load elsewhere. Those measures include reducing the batch size (not to be confused with the patch size), using a simpler model, or reducing the output size. All of those means potentially hamper training and inference. (B) Keeping a large field of view by using a small patch size of down-sampled inputs. This approach allows for a wider field of view for a constant input size while losing detail information.

To decide upon the better of the two approaches presented above, the requirements for the task at hand need to be considered. A suitable network for our task of complete distinct bone segmentation from upper-body CT scans (see 1) should provide the following: Its field of view should be sufficiently big to distinguish similar bones at different body locations, e.g. left from right humerus or the fourth from the eighth rib while keeping the computational burden in a feasible area.

The merits of high-resolution inputs—accurate details—and low-resolution inputs—a larger field of view—can be combined in many ways. Cascaded U-Nets consist of two or more individual U-Nets that are trained consecutively. A first model is trained on down-sampled input. Its one-hot encoded segmentation results are then upsampled, potentially cropped and used as additional input channels for the following model at higher resolution [15]. These approaches all have the downside of requiring the training and sequential inference of multiple models. Instead of this, we focus on end-to-end trainable models here.

Table 1 Upper-body CT dataset spatial properties

Dataset	Mean size (px)	Resolution (mm)	
		In-plane	Out-plane
Original	512 × 512× [656–1001]	0.83−0.98	1.0−1.5
Isotropic	237 × 237×403	2.0	2.0

End-to-end trained multi-resolution architectures have been proposed in histopathology whole-slide segmentation. For example, MRN [16] combines a 2D target U-Net and one context encoder with drop-skip connections crossing over at every level. MRN does not contain a context decoder or context loss and is studied on a binary segmentation problem. Another such architecture is HookNet [17], which contains both a target and a context 2D U-Net and two individual losses, but only uses skip connections in the bottleneck layer.

The purpose of our work is to address common segmentation errors that originate from a lack of global context while using 3D U-Nets for distinct bone segmentation. We propose to use a multi-resolution approach and present SneakyNet, an expansion and generalization of the MRN and HookNet architectures. We compare the segmentation accuracy, complexity, and run-time of baseline 3D U-Nets with the SneakyNet. We ablate the model components and find that the use of our generalized architecture improves the results over the HookNet and MRN variants. We will use our bone segmentation in conjunction with 3D rendering of anatomical images in augmented- and virtual reality applications, where segmentations can be used on top or in conjunction with existing transfer functions [18, 19].

Materials and methods

To assess the performance of SneakyNet on upper-body distinct bone segmentation, we train it on our in-house upper-body CT dataset. We make ablation studies on the combination of context and target information and on the optimal number of context networks.

Upper-body CT dataset

The CT images have been acquired post-mortem from body donours aged 44–103 years, 7 female and 9 male. The acquisition of the scans and the manual target segmentations have been done by specialists of the anatomical department of the University of Basel. All CT scans were taken with the body donours lying on their backs, and arms placed in front of the body. The arms are bent to various degrees, and the hands overlap in some instances.

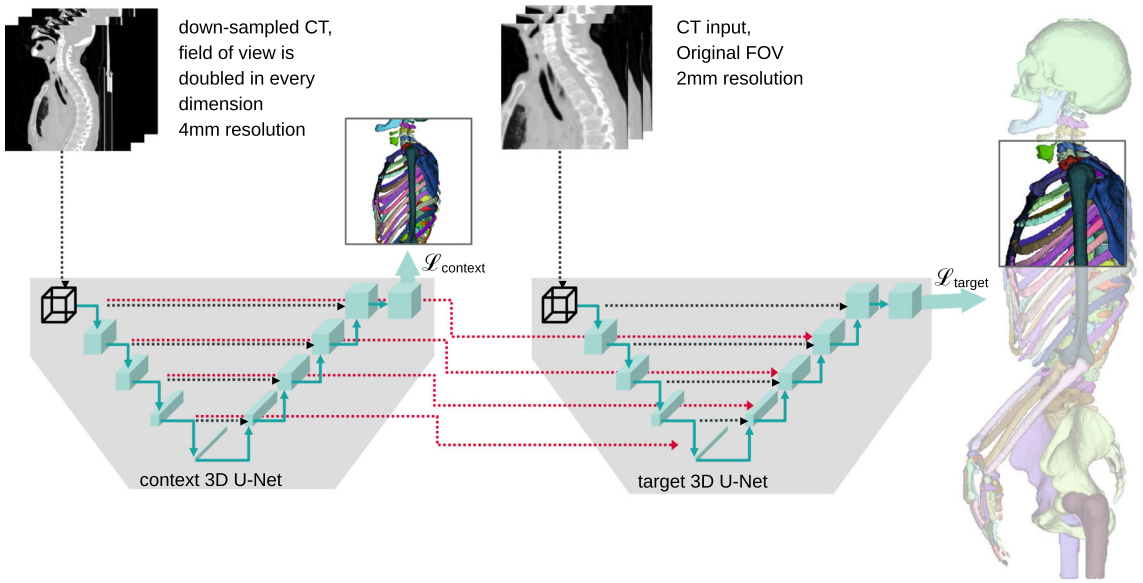

Fig. 1 Task overview: We segment 125 distinct bones from upper-body CT scans using SneakyNet, a multi-encoder–decoder network which incorporates inputs at various resolutions. The example here features one context network, but multiple are possible

Prior to using them for training and for inference, we resampled all scans to 2 mm isotropic resolution, see also Table 1. We used the same dataset, also resampled to 2 mm, in our previous publication [20].

SneakyNet architecture

In general, SneakyNet consists of one target network and one or more context networks. The target network operates on high-resolution data and eventually produces the desired segmentation maps. The context networks operate on lower resolution inputs spanning a larger field of view. Information is propagated from the context networks to the target network using crop-skip connections presented in section "Crop-skip connections". We present a detailed visual overview of the architecture with one context network in Fig. 1.

In our previous work [21], we have explored the suitability of different 2D and 3D network architectures and parameter configurations for upper-body distinct bone segmentation. We found that there is little leeway in architectural choices due to the tasks large required field of view and the many classes that are segmented in parallel. A lean 3D U-Net variant was found to work best [21]. We use this variant's architecture for our target and context U-Nets here. In our baseline computations, where we have only a target network and omit the context networks, we select the number of channels in order for our variants and the baselines to have approximately the same number of trainable parameters, to ensure that improvements not only originate from an increase in the number of trainable parameters. We use a sequence of 60, 120, 240, 480, 960 channels for our baseline 3D U-Net and 30, 60, 120, 240, 480 channels for all other networks, including the slim 3D U-Net used on patches of size 128^3, because the full network would otherwise exceed the available GPU memory. Inputs to the network are required to be multiples of 2^{M-1}, where M denotes the number of levels of the U-Net. We use the basic architecture of $M = 5$ and therefore need multiples of 16 pixels in every dimension as input.

For the target network, we use inputs of size (Sx, Sy, Sz) at full resolution. For each of the context networks, we use that input plus its surrounding area, which together span a field of view of $2^\kappa \cdot (Sx, Sy, Sz)$. We display the case of $\kappa = 1$ in Fig. 1, but also use context networks with $\kappa = 2$ and $\kappa = 3$ in our ablation studies. The context network inputs are down-sampled to reduce their size to (Sx, Sy, Sz). We perform the down-sampling using $(2^\kappa \times 2^\kappa \times 2^\kappa)$ average-pooling with a stride of 2^κ. Both target and context network inputs eventually have a size of (Sx, Sy, Sz), but at different resolutions and fields of view.

Crop-skip connections

We use crop-skip connections to transfer information from the context to the target branch. We crop the encoder output at the desired level m such that in every dimension only the central $1/2^\kappa$ part remains. This centre cube is now spatially aligned to the input of the target branch. We concatenate the centre cube to the corresponding lower level $m + \kappa$ of the target decoder to match the spatial size. We refer to the central cropping and subsequent concatenation into a lower level of the target branch as crop-skip-connection. A detailed schematic of the crop-skip connection is depicted in Fig. 2.

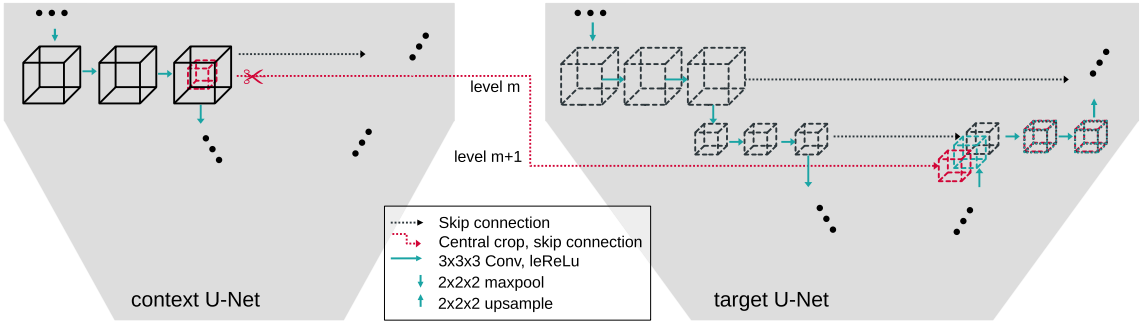

Fig. 2 Detailed view of the architecture with the first context network ($\kappa = 1$). Displayed are only two out of five levels of the U-Nets. Left: the context U-Net working on half-resolution ($1/2^{\kappa}$) data with a field of view that is double (2^{κ}) in every dimension. Right: The U-Net working with the central cropped high-resolution data. After all encoder convolutions of level m, a cropped copy of the output is skipped to the target decoder at level $m + \kappa$. The decoder receives skip connections from its own encoder and the context network[s]. The intermediate results of the decoder and all skip connections are concatenated along the channel axis before undergoing further convolutions

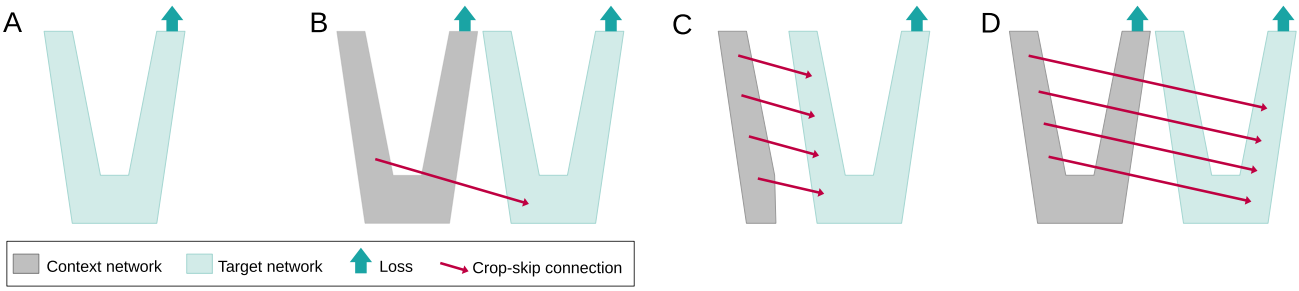

Fig. 3 Schematic of the four network configurations used in our ablation study. **A** shows a base U-Net, while **B**, **C**, and **D** show different possibilities of how to insert information into the target network, see also Sect. "Crop-skip connections" for a written description

We explore three network configurations which differ in their number of crop-skip connections and their use of a context loss, and compare it to a baseline U-Net. A visual comparison of the architectures is given in Fig. 3 and the parameters are provided in Table 2.

- *A—Baseline:* 3D U-Net with optimal configuration found for the task [21].
- *B—HookNet:* One context network with a single crop-skip connection is added to the target network. The crop-skip connection enters the target network at its bottleneck layer. This configuration is used in [17].
- *C—MRN:* Crop-skip connections connect the context encoder and the target decoder at every level. There is neither a context decoder nor a context loss function. This configuration was used in [16].
- *D—Proposed SneakyNet:* Crop-skip connections connect all levels of the context and target networks. The context network has a decoder with its own loss function.

Training and inference

Our dataset is split into 11 scans for training, 2 for validation, and 3 for testing. We use fivefold cross-validation, ensuring that every scan appears in precisely one of the cross-validation folds in the test set.

The only data augmentation we use is the patchwise sampling which doubles as random-cropping augmentation. We do not use rotation, scaling, addition of noise, or other data-augmentation steps, since these have shown to have little influence in the past for our specific problem [21]. We use a batch size of 1, combined with instance normalization. The loss is composed of an unweighted combination of the target network's loss and the losses of the K context networks. For both networks, we use the sum of the cross-entropy loss $\mathcal{L}_{\text{X-Ent}}$ and Dice-Loss $\mathcal{L}_{\text{DSC}}$ [22]. As in [21], we sum the Dice-Loss for every class separately and normalize by the number of classes. We optimized the network weights using the Adam optimizer with an initial learning rate of 0.001. We trained our networks for 100000 iterations until convergence was observed.

Our input images are padded by $(S - S_{\text{target}})/2$ all-around using edge value padding. The padding step ensures that we can sample high-resolution patch centres right to the image's border. During inference we sample patches such that the target patches overlap by 30% in every dimension to stitch the centre of the patches together. We do not apply any additional post-processing or inference-time data augmentation.

Table 2 Comparison of architectures with different field of view (FOV) of their target and context network(s)

Config	Target network FOV	Context network(s) FOV	Trainable param $\cdot 10^7$	Input pixels $\cdot 10^4$	Training time per iteration(s)
A 3D U-Net	32^3	–	5.8	3.3	0.44
	64^3	–		26.2	0.57
3D U-Net slim*	128^3	–	1.5	209.7	4.24
B HookNet	32^3	64^3	3.7	6.6	0.41
	64^3	128^3		52.4	0.72
C MRN	32^3	64^3	4.7	6.6	0.43
	64^3	128^3		52.4	1.27
D SneakyNet (ours)	32^3	64^3	4.9	6.6	0.45
		$64^3 - 128^3$	5.8	9.9	0.70
		$64^3 - 128^3 - 256^3$	6.2	13.1	3.16
	64^3	128^3	4.9	52.4	1.28
		$128^3 - 256^3$	5.8	78.6	3.11

*Operating the full 3D U-Net on patches of size 128^3 exceeds the available GPU memory

Table 3 Ablation results in DSC for different model configurations

Target patch size	32				64			
DSC	Median	σ	$-\sigma$	Nonzero DSC (%)	Median	σ	$-\sigma$	Nonzero DSC (%)
A 3D U-Net	0.64	+0.19	−0.34	94.5	0.83	+0.09	−0.27	94.5
B HookNet	0.66	+0.17	−0.34	94.1	0.85	+0.09	−0.32	95.3
C MRN	0.69	+0.16	−0.37	95.1	0.84	+0.09	−0.31	96.0
D SneakyNet (ours)	0.75	+0.14	−0.33	95.3	0.86	+0.08	−0.28	96.7

Table 4 Ablation results for the number of context networks in the SneakyNet architecture (D). Zero context networks correspond to the baseline 3D U-Nets (A) with different input patch sizes

Config	Target FOV per dim	Context FOV(s) per dim	DSC			
			Median	σ	$-\sigma$	Nonzero DSC (%)
A	32	–	0.64	+0.19	−0.34	94.5
D	32	64	0.75	+0.14	−0.33	95.3
D	32	64–128	**0.79**	+0.11	−0.33	94.4
D	32	64–128–256	0.79	+0.11	−0.33	95.9
A	64	–	0.83	+0.09	−0.27	95.6
D	64	128	**0.86**	+0.08	−0.28	96.7
D	64	128–256	0.85	+0.09	−0.28	96.1
A	128	–	0.82	+0.11	−0.30	94.3

We implemented and trained our networks using Tensorflow Keras 2.5.0. All training and inference were conducted on NVidia Quadro RTX 6000 GPUs of 24 GB RAM size.

Evaluation

We evaluate the performance of our models using a class-wise Dice Score Coefficient (DSC). To indicate the performance over all classes, we give the median and the 16 and 84 quantiles (1σ) over all classes c. To not give a distorted impression of the distribution, we exclude classes where no true positives of c have been detected and therefore $\mathrm{DSC_c} = 0$. We present the percentage of classes included as 'nonzero DSC' in Tables 3 and 4 to make up for the omission.

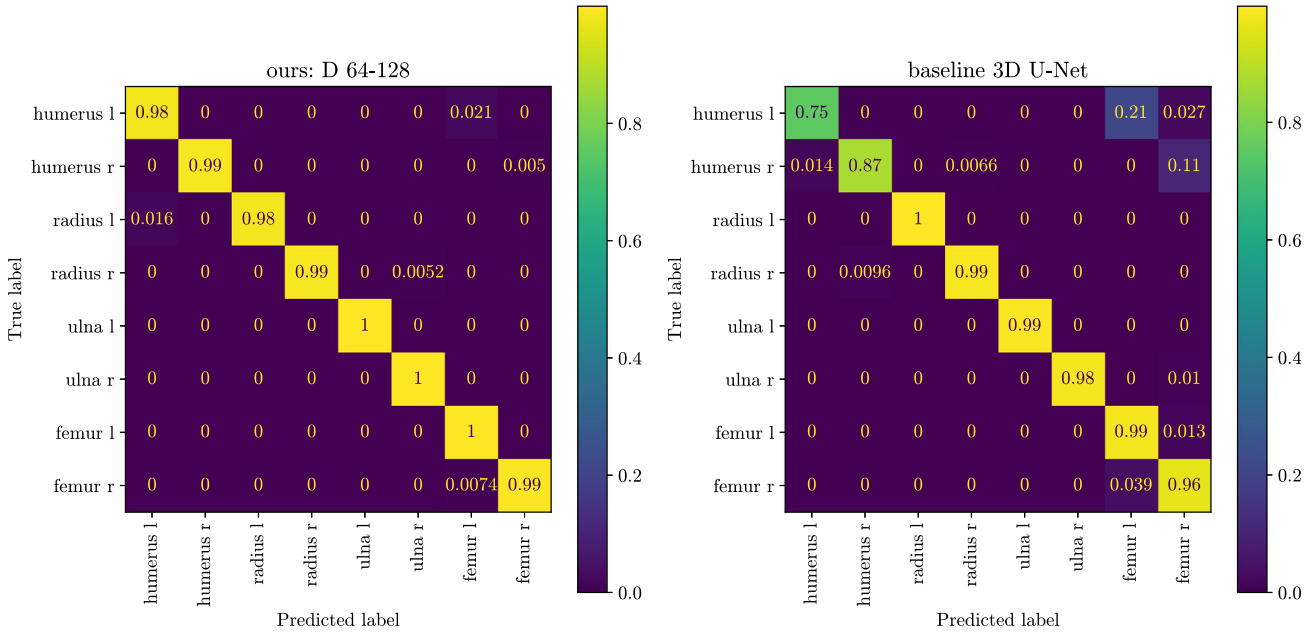

Fig. 4 Confusion matrix among the long bones of the arms and legs. With our method, there is considerably less confusion between the left and right sides of the body and between arm and leg bones

Results and discussion

Our experiments show how automated distinct bone segmentation can be improved using a multi-resolution approach. We evaluate our results on multiple target resolutions with different numbers of context networks and field of view sizes and perform an ablation study to determine the most beneficial way to combine context and target network information.

We evaluated some of the most common errors when using a baseline segmentation method. We found that the missing context information leads to similar-looking bones in different body regions being mistaken for one another. In the confusion matrix presented in Fig. 4, we observe that when using a baseline 3D U-Net, humerus pixels were predicted as femur, and the left and right humerus were confused for one another (right confusion matrix). When using context information, these errors are reduced almost entirely (left confusion matrix).

We performed an ablation study to see how different strategies of combining the context and target information within the network perform. In Table 3, we present the quantitative results. For both target patch sizes, 32 and 64, all strategies (B-D) improve upon the baseline 3D U-Net (A). The observed effect is substantially bigger when using the smaller target patch size of 32^3, where the median DCS rises from 0.64 to 0.75. The DSC still increases from 0.83 to 0.86 median DSC on the bigger target patches.

The combination of skip connections at every level and a context loss function in our proposed architecture increases the accuracy further, as compared to the HookNet [17] and the MRN [16].

In Table 4, we ablate the influence of different numbers of context networks and input patch sizes. Qualitative results are depicted in Fig. 5. Comparing the baseline 3D U-Nets with the SneakyNet results, we see that adding context networks to very small target patches of 32^3 pixels almost reaches the performance of our baseline networks operating on 64^3 patches. Going up, the SneakyNet operating on patch size 64^3 even outperforms the baseline 3D U-Net of patchsize 128^3. We recall that we had to reduce the number of channels in the baseline 128^3 network, due to memory restraints. Our ablation results suggest that the addition of context networks is more valuable in adding performance when reaching memory limits. When considering the different FOV of the context networks, we observe the best results when including context FOVs of 128^3. This covers roughly half of the L-R and A-P dimensions of the scans and seems to contain the necessary information to correctly locate bones, see, e.g. the purple lumbar vertebra in Fig. 5, which is correctly located in cases where the context FOV reaches 128^3.

We provide a comparison to other results published on distinct bone segmentation in Table 5. While a direct comparison is difficult due to different datasets, our results compare favourably to both the convolutional neural networks and shape model approach by [12], and to the hierarchical atlas segmentation by [11]. In terms of robustness, the other works are likely to cover more variation since they use larger training datasets. So far, our trained models have been evaluated

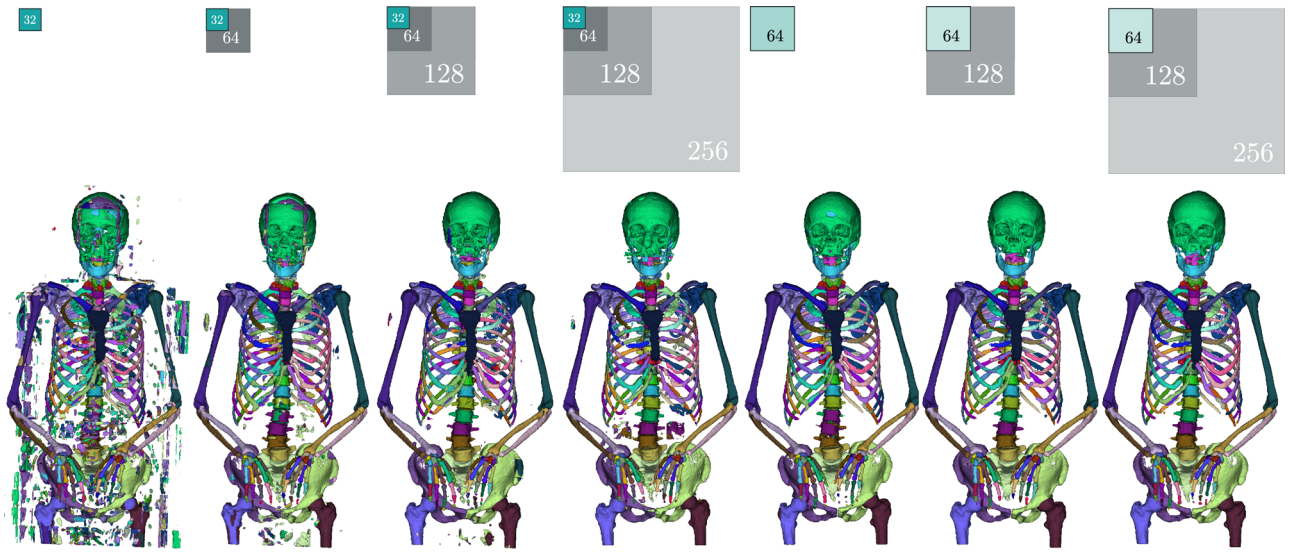

Fig. 5 Qualitative prediction results from our ablation study comparing different numbers of context networks at various resolutions. The first four results from the left were obtained using a target patch size of 32px per dimension (turquoise), and the remaining three scans with target patch sizes of 64px per dimension (light blue). The grey areas indicate the field of view of the context networks. The sizes of the squares are proportional to the prediction sizes

Table 5 Comparison of our best-performing SneakyNet (D, target patch size of 64^3 and one context network with a FOV of 128^3 pixels) to other work on distinct bone segmentation from upper-body CT. Results are in DSC

	Ours (median)	[12] (median)	[11] (mean)
L3	0.91	0.85	0.91
Sacrum	0.93	0.88	
Right 7th rib	0.78	0.84	
Clavicula	0.96		0.87
Right femur	0.97		0.92
Pelvic bones	0.96		0.86
Hamate	0.86		
Inference time per scan (min)	∼ 3		∼ 20
Scans in training dataset (#)	11	100	19
Classes (#)	125	49	62

only on data from one CT scanner and with the arm pose customary to our dataset. The use of other scanners or different poses would likely need retraining of at least parts of the model. A more in-depth analysis of how bones of different groups have performed can be found in Table 6. The segmentation performance of individual bones is affected by their size, with the small distal bones most likely to be missed or poorly segmented. For the larger, but still small carpal and metacarpal bones the segmentation performance is close to the one observed in the vertebrae.

Conclusion

This work presents improvements in distinct bone segmentation from upper-body CT. The proposed multi-resolution networks use additional inputs at a lower resolution but with a larger field of view to provide the necessary context information to assign the proper bone classes. We compared three different ways of combining the context and target information and evaluated the results using zero to three context networks. Using context networks improves the segmentation results on all target patch sizes.

Table 6 Performance of our best-performing SneakyNet (D, target patch size of 64^3 and one context network with a FOV of 128^3 pixels) on various bones and groups thereof. We first provide the four best and worst performing bones (according to their DSC values) and then the median results for all bones within given bone groups. Missed denotes the percentage of bones without any true positive pixels We first provide the four best and worst performing bones (according to their DSC values) and then the median results for all bones within given bone groups. Missed denotes the percentage of bones without any true positive pixels

Name	Size [# voxels]	DSC	Sensitivity	Precision	Missed (%)
Humerus, right	18397	0.97	0.96	0.98	0.0
Humerus, left	24061	0.97	0.96	0.98	0.0
Femur, right	26249	0.97	0.96	0.97	0.0
Pelvis, right	44601	0.96	0.96	0.97	0.0
...					
Phalanx IV, distal	38	0.33	0.28	0.52	18.8
Phalanx II, distal	36	0.24	0.18	0.56	25.0
Phalanx V, distal	26	0.18	0.10	0.50	25.0
Phalanx V, distal	18	0.15	0.13	0.32	37.5
Ribs	2688	0.80	0.76	0.87	0.0
Spine	3222	0.89	0.89	0.90	0.3
Fingers	115	0.60	0.53	0.79	11.0
Carpals and metacarpals	320	0.87	0.88	0.90	2.4
Limbs and skull	14153	0.95	0.94	0.97	0.0

Acknowledgements This work was financially supported by the Werner Siemens Foundation through the MIRACLE project. We thank Azhar Zam for valuable discussions that helped shape this work.

Funding Open access funding provided by University of Basel. This work was financially supported by the Werner Siemens Foundation through the MIRACLE project.

Data availability The upper-body CT dataset is not publicly available. An anonymized version can be shared on request.

Code availability Our code is shared at: https://gitlab.com/cian.unibas.ch/sneakynet.

Declarations

Conflict of interest None of the authors have competing interests to declare that are relevant to the content of this article.

Consent for publication Body donours signed informed consent regarding publications using their data.

Consent to participate Informed consent was obtained from all individual body donours included in the study.

Ethical approval This research study was conducted retrospectively from CT data routinely obtained from body donours. No ethical approval is required.

References

1. Deng Y, Wang L, Zhao C, Tang S, Cheng X, Deng H-W, Zhou W (2022) A deep learning-based approach to automatic proximal femur segmentation in quantitative ct images. Med Biol Eng Comput 60(5):1417–29
2. Uemura K, Otake Y, Takao M, Makino H, Soufi M, Iwasa M, Sugano N, Sato Y (2022) Development of an open-source measurement system to assess the areal bone mineral density of the proximal femur from clinical ct images. Arch Osteoporos 17(1):1–11
3. Su Z, Liu Z, Wang M, Li S, Lin L, Yuan Z, Pang S, Feng Q, Chen T, Lu H (2022) Three-dimensional reconstruction of kambin's triangle based on automated magnetic resonance image segmentation. J Orthop Res 40(12):2914–2923
4. Klein A, Warszawski J, Hillengaß J, Maier-Hein KH (2019) Automatic bone segmentation in whole-body ct images. Int J Comput Assist Radiol Surg 14(1):21–29
5. Leydon P, O'Connell M, Greene D, Curran KM (2021) Bone segmentation in contrast enhanced whole-body computed tomography. Biomed Phys Eng Exp 8(5):055010
6. Noguchi S, Nishio M, Yakami M, Nakagomi K, Togashi K (2020) Bone segmentation on whole-body ct using convolutional neural network with novel data augmentation techniques. Comput Biol Med 121:103767
7. Payer C, Stern D, Bischof H, Urschler M (2020) Coarse to fine vertebrae localization and segmentation with spatialconfiguration-net and u-net. In: VISIGRAPP (5: VISAPP), pp 124–133

8. Yang J, Gu S, Wei D, Pfister H, Ni B (2021) Ribseg dataset and strong point cloud baselines for rib segmentation from ct scans. In: International conference on medical image computing and computer-assisted intervention, pp 611–621. Springer
9. Faisal A, Khalil A, Chai HY, Lai KW (2021) X-ray carpal bone segmentation and area measurement. Multim Tools Appl 1–12
10. Wakamatsu Y, Kamiya N, Zhou X, Hara T, Fujita H (2020) Semantic segmentation of eight regions of upper and lower limb bones using 3d u-net in whole-body ct images. Nihon Hoshasen Gijutsu Gakkai Zasshi 76(11):1125–1132
11. Fu Y, Liu S, Li HH, Yang D (2017) Automatic and hierarchical segmentation of the human skeleton in CT images. Phys Med Biol 62(7):2812–2833
12. Lindgren Belal S, Sadik M, Kaboteh R, Enqvist O, Ulén J, Poulsen MH, Simonsen J, Høilund-Carlsen PF, Edenbrandt L, Trägårdh E (2019) Deep learning for segmentation of 49 selected bones in CT scans: First step in automated PET/CT-based 3D quantification of skeletal metastases. Eur J Radiol 113:89–95
13. Kamiya N, Kume M, Zheng G, Zhou X, Kato H, Chen H, Muramatsu C, Hara T, Miyoshi T, Matsuo M, Fujita H (2018) Automated recognition of erector spinae muscles and their skeletal attachment region via deep learning in torso ct images. In: International workshop on computational methods and clinical applications in musculoskeletal imaging, pp 1–10. Springer
14. Boutillon A, Borotikar B, Burdin V, Conze P-H (2022) Multi-structure bone segmentation in pediatric mr images with combined regularization from shape priors and adversarial network. Artif Intell Med 132:102364
15. Isensee F, Jaeger PF, Kohl SA, Petersen J, Maier-Hein KH (2021) nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18(2):203–211
16. Gu F, Burlutskiy N, Andersson M, Wilén LK (2018) Multi-resolution networks for semantic segmentation in whole slide images. Comput Pathol Ophthalmic Med Image Anal 11–18
17. Van Rijthoven M, Balkenhol M, Siliņa K, Van Der Laak J, Ciompi F (2021) Hooknet: Multi-resolution convolutional neural networks for semantic segmentation in histopathology whole-slide images. Med Image Anal 68:101890
18. Faludi B, Zentai N, Zelechowski M, Zam A, Rauter G, Griessen M, Cattin PC (2021) Transfer-function-independent acceleration structure for volume rendering in virtual reality. In: Proceedings of the conference on high-performance graphics, pp 1–10
19. Żelechowski M, Karnam M, Faludi B, Gerig N, Rauter G, Cattin PC (2021) Patient positioning by visualising surgical robot rotational workspace in augmented reality. Comput Methods Biomech Biomed Eng Imaging Vis 10(4):451–7
20. Schnider E, Huck A, Toranelli M, Rauter G, Müller-Gerbl M, Cattin PC (2022) Improved distinct bone segmentation from upper-body ct using binary-prediction-enhanced multi-class inference. Int J Comput Assist Radiol Surg 17(11):2113–2120
21. Schnider E, Horváth A, Rauter G, Zam A, Müller-Gerbl M, Cattin PC (2020) 3d segmentation networks for excessive numbers of classes: Distinct bone segmentation in upper bodies. In: International workshop on machine learning in medical imaging, pp 40–49. Springer
22. Milletari F, Navab N, Ahmadi S-A (2016) V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 fourth international conference on 3D vision (3DV), pp 565–571. IEEE