

## RESEARCH ARTICLE

# Efficient cascaded V-net optimization for lower extremity CT segmentation validated using bone morphology assessment

Ruurd J. A. Kuiper<sup>1,2</sup>  | Ralph J. B. Sakkers<sup>1</sup> | Marijn van Stralen<sup>2,3</sup>  |  
Vahid Arbabi<sup>1,4</sup>  | Max A. Viergever<sup>2</sup>  | Harrie Weinans<sup>1</sup>  | Peter R. Seevinck<sup>2,3</sup>

<sup>1</sup>Department of Orthopaedics, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>2</sup>Image Sciences Institute, University Medical Center Utrecht, Utrecht, The Netherlands

<sup>3</sup>MRIguidance B.V., Utrecht, The Netherlands

<sup>4</sup>Department of Mechanical Engineering, University of Birjand, Birjand, Iran

## Correspondence

Ruurd J. A. Kuiper, Image Sciences Institute, University Medical Center Utrecht, Heidelberglaan 100, Q.02.4.45, 3584 CX Utrecht, The Netherlands.

Email: [r.j.a.kuiper-4@umcutrecht.nl](mailto:r.j.a.kuiper-4@umcutrecht.nl) and [ruurdkuiper@gmail.com](mailto:ruurdkuiper@gmail.com)

## Funding information

Nederlandse Organisatie voor Toegepast Natuurwetenschappelijk Onderzoek, Grant/Award Number: 15479

## Abstract

Semantic segmentation of bone from lower extremity computerized tomography (CT) scans can improve and accelerate the visualization, diagnosis, and surgical planning in orthopaedics. However, the large field of view of these scans makes automatic segmentation using deep learning based methods challenging, slow and graphical processing unit (GPU) memory intensive. We investigated methods to more efficiently represent anatomical context for accurate and fast segmentation and compared these with state-of-the-art methodology. Six lower extremity bones from patients of two different datasets were manually segmented from CT scans, and used to train and optimize a cascaded deep learning approach. We varied the number of resolution levels, receptive fields, patch sizes, and number of V-net blocks. The best performing network used a multi-stage, cascaded V-net approach with  $128^3-64^3-32^3$  voxel patches as input. The average Dice coefficient over all bones was  $0.98 \pm 0.01$ , the mean surface distance was  $0.26 \pm 0.12$  mm and the 95th percentile Hausdorff distance  $0.65 \pm 0.28$  mm. This was a significant improvement over the results of the state-of-the-art nnU-net, with only approximately 1/12th of training time, 1/3th of inference time and 1/4th of GPU memory required. Comparison of the morphometric measurements performed on automatic and manual segmentations showed good correlation (Intraclass Correlation Coefficient [ICC]  $>0.8$ ) for the alpha angle and excellent correlation (ICC  $>0.95$ ) for the hip-knee-ankle angle, femoral inclination, femoral version, acetabular version, Lateral Centre-Edge angle, acetabular coverage. The segmentations were generally of sufficient quality for the tested clinical applications and were performed accurately and quickly compared to state-of-the-art methodology from the literature.

## KEYWORDS

bone, diagnostic Tmaging, hip, knee

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Orthopaedic Research*® published by Wiley Periodicals LLC on behalf of Orthopaedic Research Society.

## 1 | INTRODUCTION

Differentiating between different tissue types on radiological scans, also known as segmentation, is an important part of modern medical image analysis. In orthopaedics, it is most often used to differentiate between bones, muscle, ligaments and cartilage on X-ray, computed tomography (CT) and magnetic resonance imaging (MRI) scans.<sup>1</sup> Segmentation of bones of the lower extremities has a multitude of use-cases in the clinical practice, such as the measurement of hip, ankle and knee joint range of motion,<sup>2,3</sup> determination of anatomical and mechanical axes of the full lower limbs,<sup>4-6</sup> indication of landmarks,<sup>7</sup> and the fabrication of surgical tools<sup>8</sup> and medical implants.<sup>9</sup> However, manual segmentation is a time consuming task that requires knowledge and expertise of dedicated software.<sup>1</sup> Difficulties encountered in bone segmentation include differing image intensity values between cortical and cancellous bone, insufficient resolution to differentiate between bones in joint spaces and low signal-to-noise ratios in scans. Therefore, automatic segmentation methods have been developed that aim to accelerate the segmentation process and remove interobserver variability.

Initial (semi-) automatic segmentation algorithms relied mostly on intensity-based methods, such as thresholding, region growing, or edge detection. Later, more sophisticated methods such as Statistical Shape Models (SSMs) and Atlas-based methods were applied.<sup>10</sup> For example, Audenaert et al.<sup>11</sup> proposed a SSM based pipeline for semantic segmentation of lower body CT scans. Recently, deep learning networks, and more specifically Convolutional Neural Networks (CNNs), have shown to be an effective approach for CT bone segmentation.<sup>9,12-15</sup> An often used network architecture for medical image segmentation is the U-net,<sup>16</sup> which can be extended to 3D and is then called 3D U-net<sup>17</sup> or V-net.<sup>18</sup> Noguchi et al.<sup>19</sup> showed that the V-net architecture is suitable for binary segmentation of the complete human skeleton from CT.

The lower extremity CT scans used in this study have a large field of view and comparatively high resolution, resulting in a large amount of data to be processed. An issue encountered in CNNs when segmenting large volumes is the increase in graphical processing unit (GPU) memory that is needed when the input for the network becomes larger. Patch-based networks circumvent this problem by segmenting only part of the image at a time. This reduces the amount of voxels the network needs to process each iteration, but also reduces the amount of contextual information the network has for each prediction. Additionally, due to the large number of patches per scan inference can become prohibitively slow. Multiple studies have tried to balance the amount of information available to the network and its memory consumption by implementing cascaded approaches<sup>20-23</sup> or by using dilated (atrous) convolutional layers.<sup>24,25</sup> Isensee et al.<sup>26</sup> introduced a (cascaded) U-net based network called nnU-net that automatically adapts its architecture and processing steps to the dataset on which it is used. However, none so far have studied the influence of varying the number of cascades by more than two. Additionally, full lower extremity scans have a larger number of voxels than usually encountered in these

studies, increasing the need for lightweight method for efficient segmentation.

Two examples of the clinical application of segmentations of the lower extremities include the hip-knee-ankle angle (HKA) measurement and hip morphometry assessment. Although HKA measurements were conventionally performed manually on 2D radiographs,<sup>27</sup> recent studies in knee realignment planning show an increased trend towards 3D CT planning and 3D measurements.<sup>28</sup> Hip morphometry assessment was conventionally also performed on X-rays or CT scans by manual segmentation and indication of landmarks.<sup>29-32</sup> Recently, Palit et al.<sup>2</sup> performed automatic assessment of the location of bone impingement and range of motion analysis using manual segmentations of the bone from CT. Lerch et al.<sup>3</sup> showed that MR-based segmentations of the hip joint could also be used. Zeng et al.<sup>33</sup> then showed that automatic MRI segmentation also correlated well with manual CT segmentation, removing the need for time-consuming manual segmentations.

In this study, we propose a deep learning based approach for automatic, accurate, fast and memory efficient segmentation of bones from the entire lower extremities. For this purpose, we introduce a new lightweight variant of the cascaded approach for the V-net architecture to reduce memory requirements and speed up training and inference. Additionally, we applied a sampling scheme that maximizes use of the available information. This approach was evaluated in four steps:

1. The optimal network architecture was studied by comparing the results with differing number of cascades and input configurations of the cascaded V-net.
2. The best performing network was trained as a fivefold cross-validation and the results were compared to existing methods found in the literature and the recent nnU-net.<sup>26</sup>
3. The clinical applicability was evaluated by performing two different medical assessments using both the manual and automatic segmentations:
  - a. The hip morphology was measured using proprietary software developed by the Move Forward service (Clinical Graphics, Zimmer Biomet).
  - b. The HKA angle was measured using an automatic workflow based on the methods from Frnstahl et al.<sup>4</sup>
4. The best performing network was evaluated on a dataset with different subject demographics and acquisition parameters to evaluate the robustness of the method.

## 2 | METHODS

### 2.1 | Data

#### 2.1.1 | University Medical Center Utrecht (UMCU) dataset

For the training and initial evaluation of the networks, fifty CT scans of the lower extremity were used. The anonymized data were

acquired retrospectively and was judged not to be subject to the Medical Research Involving Human Subjects Act (WMO) by the Medical Ethical Committee, as described in IRB Protocol Number 16-612/C. Lower extremity CT scans of patients who had undergone CT scanning due to unrelated medical reasons (i.e., vascular indications) were collected from the UMCU (Utrecht, The Netherlands) and anonymized. The mean age of the male patients was 61 years (*SD*: 10 years) and of the female patients 53 years (*SD*: 15 years).

All CT scans were acquired with either the Philips iCT scanner or Philips Brilliance 64 (Philips Medical Systems, Best, The Netherlands). The CT acquisition parameters were: tube voltage = 120 kVp, tube current = 31–347 mA, effective dose = 35–150 mAs, slice thickness = 1 mm, slice increment = 0.7 mm, pixel spacing = 0.63–0.98 mm, matrix size = 512 × 512 pixels. The iDOSE4 reconstruction algorithm was used.

### 2.1.2 | New Mexico Decedent Image Database (NMDID) dataset

To evaluate the robustness of the final deep learning segmentation network, a dataset of over 15,000 deceased subjects with different acquisition parameters and population distribution was acquired from the NMDID.<sup>34</sup> We selected 10 patients with death by natural cause, age under 50 and similar positioning to patients from the UMCU dataset: 6 male and 4 female patients, with a mean age of 34 years (*SD*: 6.4 years). For each subject, two CT scans were available which included the lower extremities. A torso scan, including the coxae and femur, and a lower leg scan, including the tibia, fibula, talus, calcaneus and femur. All CT scans were acquired on a Philips Brilliance Big Bore scanner (Philips Medical Systems). The CT acquisition parameters were: tube voltage = 120 kVp, tube current = 82–245 mA, effective dose = 100–301 mAs, slice thickness = 1 mm, slice increment = 0.5 mm, pixel spacing = 0.63–1.17 mm, matrix size = 512 × 512 pixels.

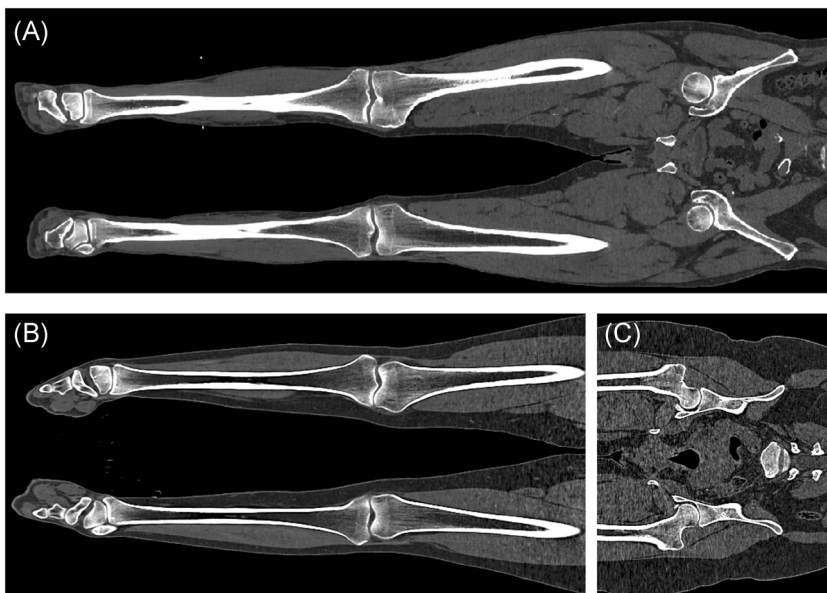
The images were resampled using trilinear interpolation to isotropic 0.8 × 0.8 × 0.8 mm to match the voxel size of the UMCU dataset more closely. Due to high noise levels in the proximal part of the lower leg scan and in the torso scan, a Gaussian filter ( $\sigma = 0.5$ , kernel size = 4) was applied before segmentation. Examples of CT scans from both datasets are shown in Figure 1.

### 2.1.3 | Segmentation

For both datasets the bones were segmented semantically, i.e., each bone was given an individual label. The tibia, fibula, talus, calcaneus, femur, and coxae were manually segmented using a combination of the CT Bone Segmentation Module and manual editing in Mimics (Mimics Medical 21.0, Materialise). Two operators with 2 years of experience using Mimics segmented the UMCU dataset. Both operators checked each segmentation for accuracy. A single operator with 4 years of experience using Mimics segmented the NMDID dataset. Operators reported average segmentation times of up to 20 min per label for each patient.

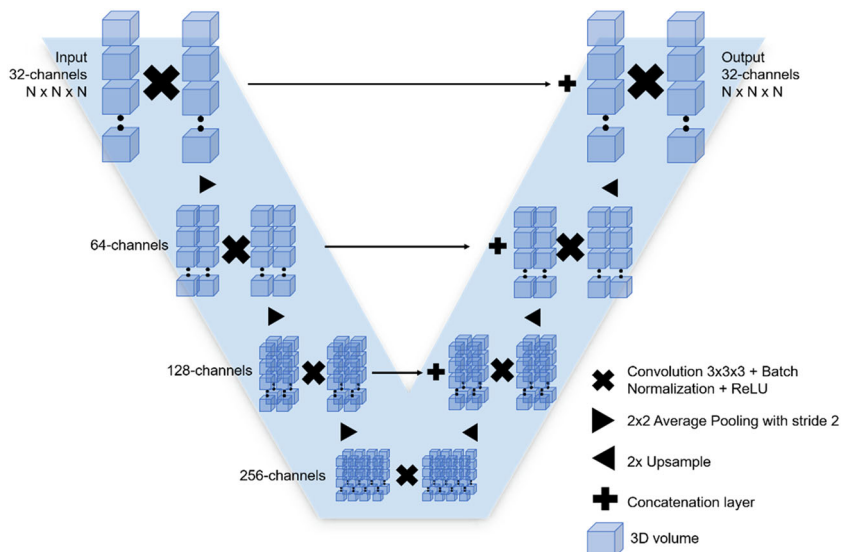
## 2.2 | V-Net

The network that was used was based on the 3D U-net<sup>17</sup> and V-net architecture,<sup>18</sup> which are 3D extensions of the original U-net.<sup>16</sup> In the remainder of the paper, we refer to these 3D network architectures as V-net. To reduce the memory requirements of the network, we used a lightweight implementation with only four stages per V-net, and only a single convolutional block per stage. Each convolutional block consists of a 3 × 3 × 3 convolution, followed by a batch normalization layer and a Rectified Linear Unit (ReLU) activation layer. At the end of the V-net a final convolutional layer is appended with a voxel-wise softmax layer to transform the output to a



**FIGURE 1** Examples of the CT scans of two patients. (A) UMCU dataset, lower leg scan. (B) NMDID dataset, lower leg scan. (C) NMDID dataset, torso scan, cropped to include only the femora and coxae. CT, computerized tomography; NMDID, New Mexico Decedent Image Database; UMCU, University Medical Center Utrecht

**FIGURE 2** Overview of the light-weight version of the V-net<sup>18</sup> architecture that was used in this study. An input patch with shape  $N \times N \times N$  was used as input. The network consists of a downsampling (left side) and upsampling (right side) portion, where the resolution decreases and increases, respectively. In each level of the network a single convolutional block was used, of which the number of channels doubled for each level of the network



probabilistic segmentation of the input patch. An overview of this implementation of the V-net architecture is shown in Figure 2.

### 2.3 | Cascaded networks

The lower extremity CT-scans have a high resolution and large spatial FOV. To segment the bones semantically from these scans a deep network with a large receptive field would be preferred. However, a single V-net with these attributes would necessitate high amounts of GPU memory and could lead to slow inference. A cascaded lightweight V-net strategy was therefore employed, which reduces the amount of input voxels while maintaining a large receptive field. This could speed up inference while minimizing the memory constraints on the GPU.

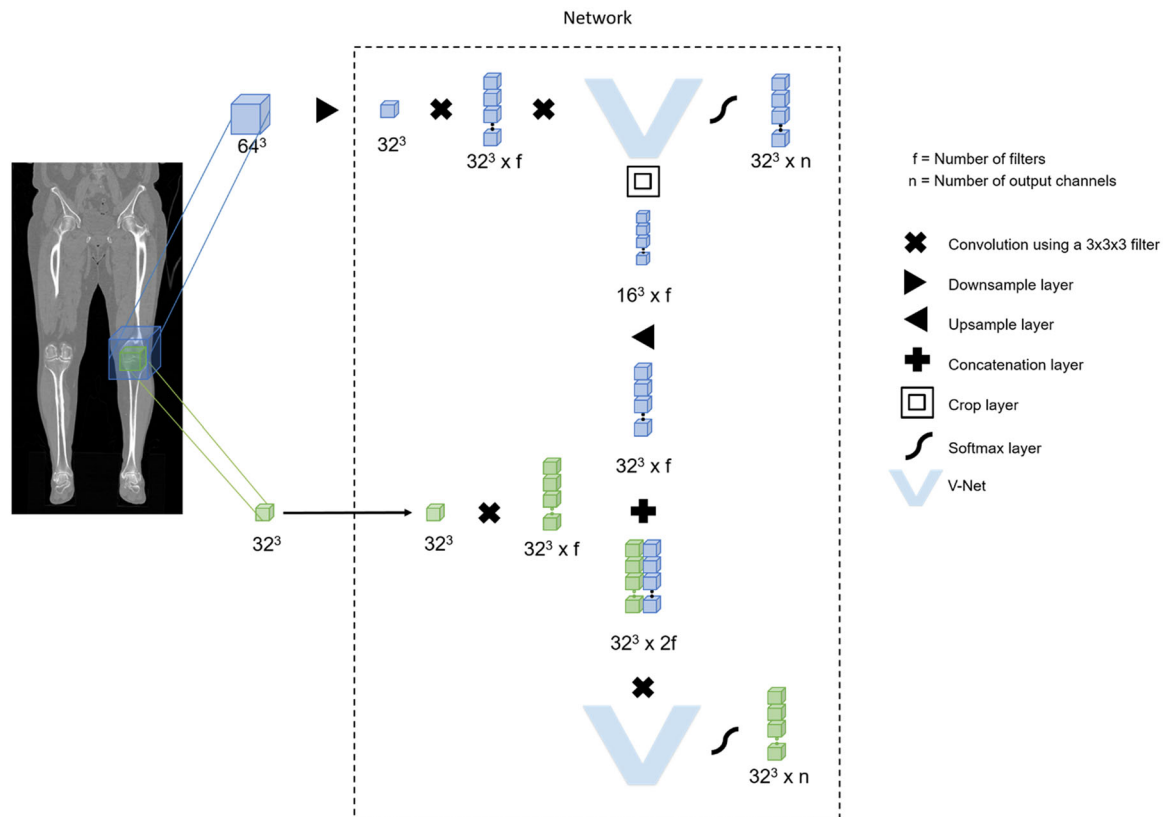
The cascaded strategy used multiple consecutive V-nets. The first V-net took as input a large FOV patch from the CT that had been down-sampled, such that information from a large receptive field was available, albeit at a lower resolution. The prediction of this patch was then used by the consecutive V-net, which also used an additional CT patch at a higher resolution, but with a smaller FOV. This continued until a full resolution patch was used as input to the final V-net, after which the final full resolution patch was predicted. The network outputs a semantic segmentation prediction for each V-net, which was used during training, but only the prediction at the full resolution was used during inference. This strategy allowed the network to process information from a large FOV while maintaining low memory requirements. Figure 3 shows an example of a two-stage cascaded V-net. All network configurations that were studied and their metrics are summarized in Table 1. The network was implemented in such a way that it automatically adapted its architecture to the number and sizes of the input patches.

All patches were downsampled using trilinear interpolation to the size of the smallest patch before being put into the network. For example, a  $128^3-64^3-32^3$  configuration denotes a network that takes a

$128 \times 128 \times 128$  patch, a  $64 \times 64 \times 64$  patch and  $32 \times 32 \times 32$  patch as input, all downsampled to  $32 \times 32 \times 32$  voxels. This configuration thus has a receptive field twice larger than the V-net proposed by Çiçek et al.,<sup>17</sup> while the number of voxels used as input to the network is ten times smaller, as shown in Table 1. To make optimal use of the cascaded V-net strategy an inference sampling strategy was used that performed inference on all smaller patch, instead of only the center patch. A detailed description of the training and sampling strategy is given in Supporting Information Material: Appendix A.

### 2.4 | Study design

The study design was divided into four stages. In stage 1, the number of cascaded networks and the input patch sizes were varied as summarized Table 1 to find the optimal input and network configuration. The 50 CT scans of the UMCU dataset were divided into sets of 35 scans for training, 5 scans for validation and 10 scans for testing. In stage 2, the configuration of the best performing network of stage 1 was trained such that all scans could be automatically segmented. In a fivefold cross-validation method, 40 scans were used for training and 10 for testing in each fold, as well as training of the cascaded version of nnU-net as a benchmark to compare our results. All training and inference parameters, settings and pre- and postprocessing were kept the same as described in the study by Isensee et al.,<sup>26</sup> of which the code is available at GitHub (<https://github.com/MIC-DKFZ/nnUNet>). In stage 3, the results of the proposed network from stage 2 were used to compare the outcome of automatic HKA measurement and hip morphometry assessment between manual and automatic segmented scans. In Stage 4, the best performing network was trained on all 50 scans of the UMCU dataset using the same settings as for the five-fold cross-validation, and then evaluated on the NMDID dataset to test the robustness of the network to different subject demographics and acquisition parameters.



**FIGURE 3** Overview of an example of the cascaded V-net architecture. In this example, two patches were chosen of  $64^3$  and  $32^3$  voxels, the smaller patch from within the larger patch. First, at the top of the network the  $64^3$  patch was downsampled to the same size as the smaller patch. The lightweight V-net shown in Figure 1 was then used to segment the patch. The output was both converted to a probability map for each segmentation using the Softmax-layer (on the right), as well as cropped to be used as additional information for the smaller, high resolution patch (downward). Next, the output of the first V-net was cropped and concatenated to the  $32^3$  full resolution patch. This was then used as input for the second V-net. After applying the Softmax-layer, the  $32^3$  full resolution probability prediction for each segmentation is the final output of the network

**TABLE 1** Summary of the different configurations of the cascaded V-net architecture that were compared

Network	Nr. of V-nets	Receptive field	Nr. of input voxels	Nr. of parameters	Additional GPU memory per patch (MB)
Proposed Cascaded V-net	1	$32^3$	$3.3 \times 10^4$	$1.2 \times 10^6$	51
	1	$64^3$	$2.6 \times 10^5$	$1.2 \times 10^6$	419
	1	$128^3$	$2.1 \times 10^6$	$1.2 \times 10^6$	3340
	2	$64^3-32^3$	$6.6 \times 10^4$	$2.4 \times 10^6$	52
	2	$128^3-32^3$	$6.6 \times 10^4$	$2.4 \times 10^6$	52
	2	$128^3-64^3$	$5.2 \times 10^5$	$2.4 \times 10^6$	421
	3	$128^3-64^3-32^3$	$9.8 \times 10^4$	$3.6 \times 10^6$	53
4	$256^3-128^3-64^3-32^3$	$1.3 \times 10^5$	$4.8 \times 10^6$	58	
3D U-net <sup>17</sup>	1	$132 \times 132 \times 116$	$2.0 \times 10^6$	$1.9 \times 10^7$	— <sup>a</sup>
V-net <sup>18</sup>	1	$128 \times 128 \times 64$	$1.0 \times 10^6$	$6.6 \times 10^7$	— <sup>a</sup>

Note: The receptive field denotes the shape and size in voxels, that is used for each stage of the V-net. The number of input voxels, number of trainable deep learning parameters and GPU memory consumed during inference are all a consequence of the number of V-nets and size of the receptive field. The cascaded V-net GPU memory had a base memory consumption of approximately 685 MB for each configuration, in the table the additional GPU memory used per patch during inference is shown. The 3D U-net and V-net were included as comparison.

Abbreviation: GPU, graphical processing unit.

<sup>a</sup>Not available in the study.

## 2.5 | Evaluation

### 2.5.1 | Segmentation analysis

Outcomes of the automatic segmentations were compared to the manual segmentations based on four commonly used metrics<sup>35</sup>: the Dice Similarity Coefficient (DSC), symmetric Mean Absolute Surface Distance (MASD), Hausdorff Distance (HD) and the 95th percentile of the Hausdorff Distance (HD95).

### 2.5.2 | HKA and hip morphology assessment

We used both the manual and the automatic segmentations from the five-fold cross-validation to measure the HKA and the morphology of the hip joint for use in clinical practice.

The HKA was determined from the segmentations using an adapted version of the original protocol developed by Fürnstahl et al.<sup>4</sup> such that no manual steps were necessary. This ensured that any differences in the measurements were due to differences between the manual and automatic segmentations, and not due to intra- or interrater variability. The adjusted protocol is described in-depth in Supporting Information Material: Appendix B.

A commercially available software tool, the Move Forward service (Clinical Graphics, Zimmer Biomet) was used to calculate the hip morphometry as defined by six different parameters: alpha angle, femoral inclination, femoral version, acetabular version, Lateral Centre-Edge (LCE) angle and acetabular coverage. An experienced user manually placed the five required landmarks medial and lateral femoral condyles, medial and

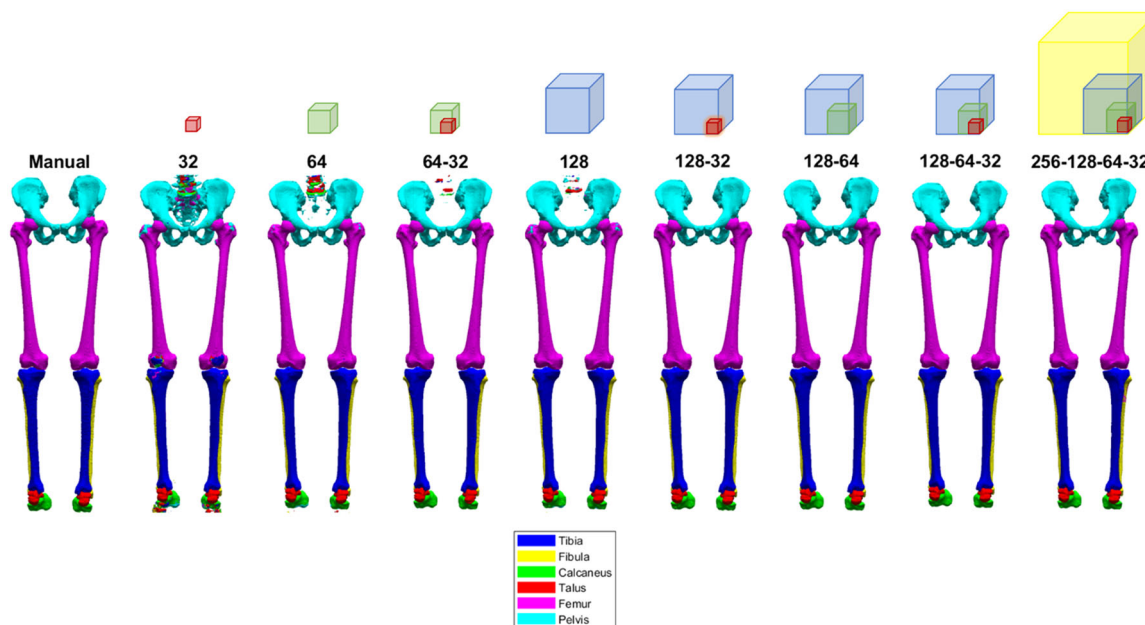
lateral femoral epicondyles and the pubic tubercle by indicating the points on three orthogonal slices of the CT. As the landmarks were placed on the CT image and not on the segmentation, the landmarks remained the same for both the manual and automatic method. A visual representation of the measurement of each of the parameters is shown in Appendix C.

## 3 | RESULTS

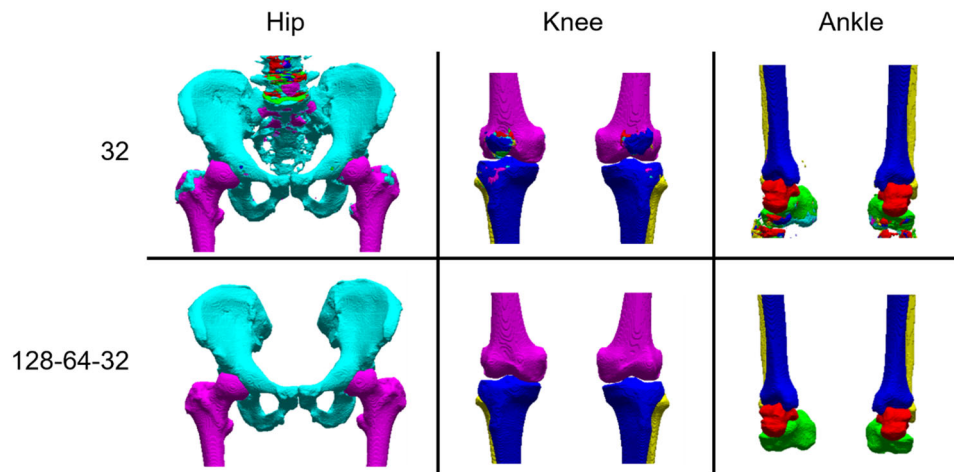
### 3.1 | Cascaded V-net optimization

The 3D models generated from the automatic segmentations visualized in Figure 4 show the different segmentation results for one subject. The detailed comparison in Figure 5 shows that most segmentation errors appear close to the joints, and in places where unlabelled bones were present such as the spine, patella and small bones in the feet. This example showed that a small receptive field of the network negatively affected its ability to discern between bones. The DSC, MASD, HD95, and HD reported for each network configuration in Table 2 reflect these observations. One-way analysis of variance analysis was performed and showed that only the network with a  $32^3$  voxel input was found to have a statistically different mean from the other configurations for any of the metrics (DSC:  $p = 8 \times 10^{-7}$ , MASD:  $p = 3.2 \times 10^{-16}$ , HD95:  $p = 3.5 \times 10^{-17}$ , HD:  $p = 5.2 \times 10^{-12}$ ).

The distribution of the segmentation metrics for each bone in Figure 6 show that the networks generally performed better in the larger and wider bones such as the femur and tibia, and worse in the thinner bones such as the fibula or coxae. As can be seen from



**FIGURE 4** 3D visualization of the segmentation result of one subject using each input configuration, before postprocessing. The cubes in the top row show the relative patch sizes, with  $32^3$  (red),  $64^3$  (green)  $128^3$  (blue) and  $128^3$  (yellow) voxels. The cubes are not to scale relative to the segmentations



**FIGURE 5** Detailed view of the differences in segmentation before postprocessing when using a small,  $32^3$ -voxel receptive field and a larger,  $128^3$ -voxel receptive field

Input	Dice		MASD		HD95		HD	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
32	0.958	0.024	0.97	1.31	5.64	11.04	19.5	20.2
64	0.969	0.014	0.41	0.13	1.15	0.91	11.2	8.5
128	0.967	0.015	0.46	0.17	1.38	1.37	12.7	8.2
64-32	0.968	0.014	0.40	0.09	1.00	0.26	8.1	9.7
128-32	0.970	0.014	0.37	0.05	0.90	0.12	5.5	7.1
128-64	0.970	0.015	0.38	0.10	0.95	0.33	5.9	10.0
128-64-32	<b>0.971</b>	0.013	<b>0.36</b>	0.06	<b>0.87</b>	0.12	<b>5.3</b>	8.0
256-128-64-32	0.969	0.014	0.40	0.11	0.98	0.45	7.2	10.4

**TABLE 2** Image analysis metrics for the different input configurations of the network after training

Note: The best performing input configuration is shown in bold for each metric.

Abbreviations: HD, Hausdorff Distance; MASD, Mean Absolute Surface Distance.

the table and the figure, the input configuration with  $128^3$ - $64^3$ - $32^3$  voxel FOVs attained the highest DSC and lowest MASD, HD95 and HD, indicative of best performance. This input configuration was used to train the network in fivefold cross validation, to provide automatic segmentations for the final segmentation and hip morphometry assessment.

### 3.2 | Fivefold cross-validation

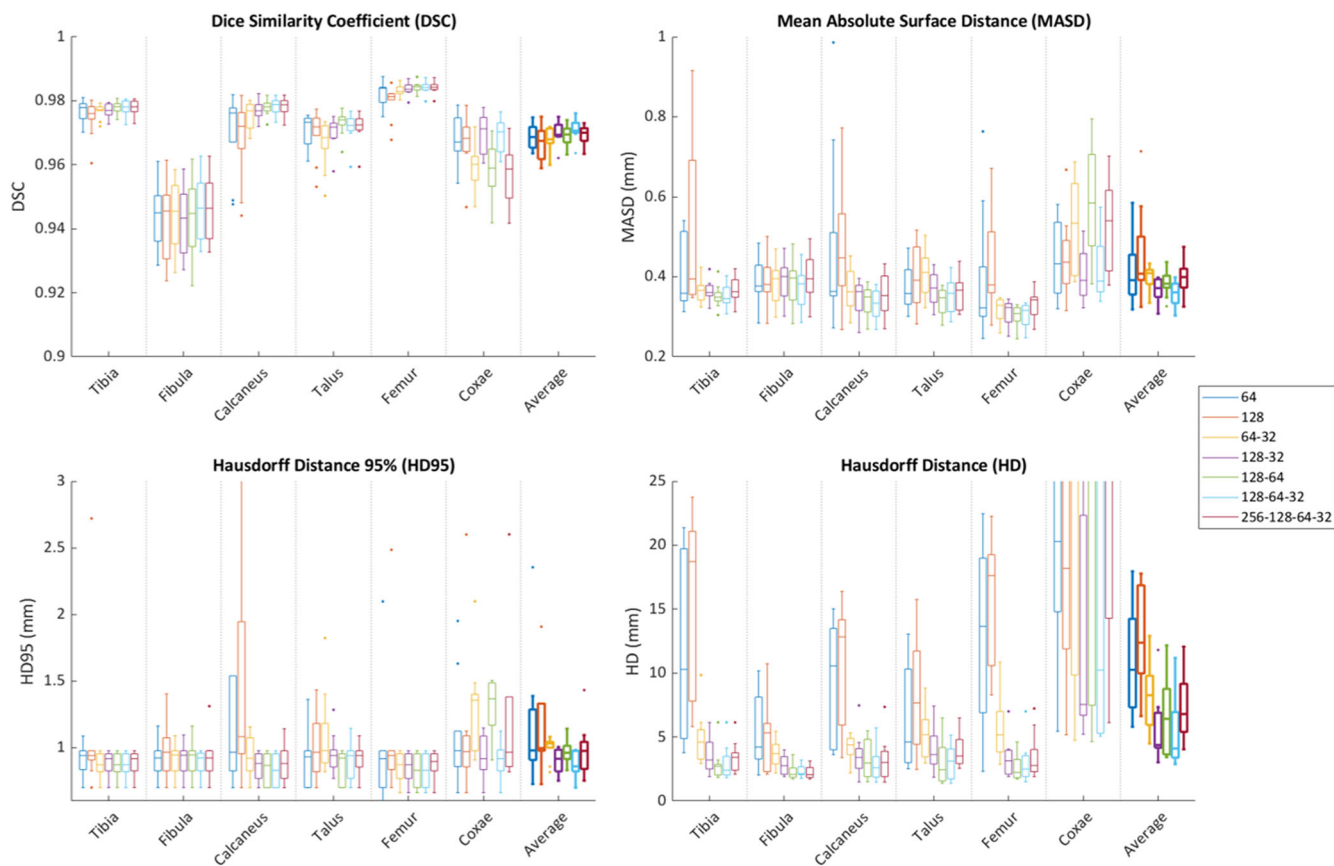
The results of the fivefold cross validation were compared to the best results of other automatic segmentation methods for which the MASD or HD was reported in the literature.<sup>11,33,36-44</sup> The proposed method achieved lower MASD than other methods for most bones except the coxae (Table 3). The MASD was also lower for most bones when compared to the interobserver MASD for manual segmentations that were reported by Audenaert et al.<sup>11</sup> Furthermore, The HD of the proposed method was also lower than those reported in the literature, in general, except for the coxae and talus. The average

DSC over all bones was  $0.98 \pm 0.01$  and the HD95 was  $0.65 \pm 0.13$  mm.

### 3.3 | Comparison to nnU-net

Figure 7 shows the results of the comparison of proposed cascaded network and the cascaded version of nnU-net run on the UMCU dataset. The two-tailed paired *t* test was performed for each metric to check for significant differences between the average results of nnU-net and the proposed method. Bonferroni correction was applied to compensate for multiple testing, resulting in  $p < 0.0125$  to be considered significant. The proposed method performed significantly better than nnU-net for the mean DSC ( $p = 9 \times 10^{-4}$ ), MASD ( $p = 7.8 \times 10^{-5}$ ) and HD95 ( $p = 0.0048$ ). For the HD no significant difference was found ( $p = 0.34$ ).

The two stage cascaded nnU-net necessitated separate training for both stages of the network. Each iteration took on average 514 s, for an average of 143 h of training for each of the fivefolds on our



**FIGURE 6** Segmentation similarity metrics calculated on the automatic segmentation versus the manual segmentation for each bone, for each network architectures. Average values indicate the metric averaged over the bones. The 32<sup>3</sup> input configuration was left out as its results deviated too much to be able to represent them clearly in the figure

**TABLE 3** Comparison of the MASD and HD between the automatic and segmentations performed on the UMCU and NMDID datasets with the best automatic segmentation methods found in the literature

	Proposed method UMCU dataset		Proposed method NMDID dataset		Literature		Literature interobserver <sup>11</sup>	
	MASD (mm)	HD (mm)	MASD (mm)	HD (mm)	MASD (mm)	HD (mm)	MASD (mm)	HD (mm)
Tibia	<b>0.23 ± 0.04</b>	2.91 ± 1.53	0.48 ± 0.05	5.08 ± 2.36	0.63 ± 0.11 <sup>11</sup>	4.07 ± 2.15 <sup>11</sup>	0.39 ± 0.19	1.88 ± 0.47
Fibula	<b>0.25 ± 0.05</b>	2.04 ± 1.59	0.43 ± 0.05	3.96 ± 2.17	0.76 ± 0.76 <sup>11</sup>	3.76 ± 1.17 <sup>11</sup>	0.61 ± 0.08	2.25 ± 0.71
Calcaneus	<b>0.21 ± 0.03</b>	2.73 ± 1.32	0.60 ± 0.06	5.48 ± 1.99	0.53 ± 0.16 <sup>11</sup>	2.90 ± 0.77 <sup>11,17</sup>	0.40 ± 0.12	1.67 ± 0.34
Talus	<b>0.23 ± 0.03</b>	3.00 ± 1.95	0.63 ± 0.11	5.47 ± 3.16	0.57 ± 0.12 <sup>11</sup>	2.97 ± 0.59 <sup>11</sup>	0.44 ± 0.08	2.21 ± 0.41
Femur	<b>0.14 ± 0.06</b>	3.54 ± 5.54	0.58 ± 0.07 <sup>a</sup>	5.03 ± 3.20 <sup>a</sup>	0.2 ± 0.1 <sup>33</sup>	4.34 ± 0.86 <sup>39</sup>	0.41 ± 0.15	2.30 ± 0.98
Coxae	0.31 ± 0.10	9.22 ± 8.56	0.56 ± 0.13	9.17 ± 5.43	0.30 ± 0.5 <sup>33</sup>	5.5 <sup>41,b</sup>	0.41 ± 0.20	3.74 ± 2.68

Note: Values are shown as mean ± standard deviation. Values performing better than found in the literature are shown in bold.

Abbreviations: CT, computerized tomography; HD, Hausdorff Distance; MASD, Mean Absolute Surface Distance; NMDID, New Mexico Decedent Image Database; UMCU, University Medical Center Utrecht.

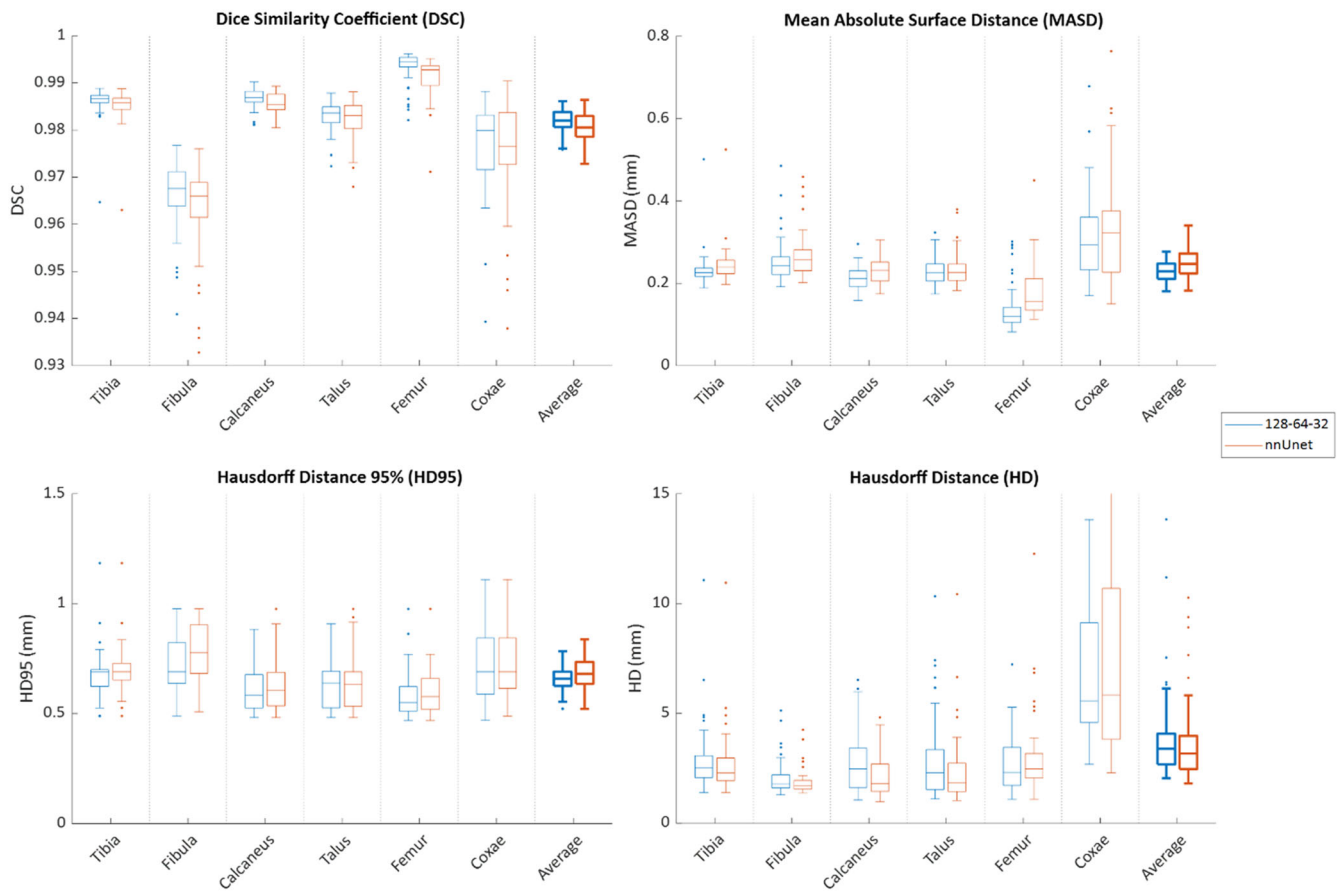
<sup>a</sup>Mean and standard deviation of combined torso and lower leg CT.

<sup>b</sup>Standard deviation unknown.

system, resulting in a total of approximately 60 days of GPU time. The training of the proposed network on the same system took only 24 h per fold, and as all stages of the cascaded V-net were imbedded in one network, the total training for all folds took approximately 5

days of GPU time. Inference using nnU-net took on average 55 min per scan, whereas the proposed network only needed 20 min per scan. For inference, nnU-net requires at least 4GB of GPU memory, while the proposed approach requires less than 1GB.





**FIGURE 7** Segmentation similarity metrics calculated on the fivefold cross-validated automatic segmentation versus the manual segmentation for each bone, for the 128–64–32 implementation of the proposed network, compared with nnU-net

### 3.4 | HKA measurement

The agreement between measurements on manual and automatic segmentations was calculated using the two-way mixed effects, absolute agreement, single measurement Intraclass Correlation Coefficient, or ICC(3,1).<sup>45,46</sup> For the HKA, the ICC was 0.976, indicating excellent reliability (ICC > 0.90), and the mean absolute difference (MAD) was  $0.18 \pm 0.36^\circ$ . These values are comparable with the interrater agreements reported by Jud et al.,<sup>28</sup> who reported an ICC of 0.988 and a MAD of  $0.4 \pm 0.5^\circ$ . The difference in landmark placement in the hip was  $0.19 \pm 0.16$  mm, in the knee  $1.2 \pm 1.6$  mm and in the ankle  $0.50 \pm 0.49$  mm.

### 3.5 | Hip morphology assessment

Table 4 shows the mean and standard deviations of all hip morphometric parameters compared to the mean and standard deviation of each parameter as they are found in the healthy adult population according to Toogood et al.<sup>32</sup> (alpha angle, femoral neck inclination, femoral version), Dandachli et al.<sup>29</sup> (acetabular coverage), Tannast et al.<sup>31</sup> (LCE angle) and Hingsammer et al.<sup>47</sup> (acetabular version).

Comparison shows that the results correspond closely to the healthy population, for both the manual and automatic segmentation.

In Table 5 the mean and MAD of all hip parameters are shown. The ICC(3,1) for the alpha angle was 0.822 indicating good reliability ( $0.75 < \text{ICC} < 0.90$ ), while the other parameters had an ICC of over 0.967, indicating excellent reliability (ICC > 0.90).<sup>49</sup> Furthermore, in Table 5 the differences were compared to the results of the studies of Zeng et al.<sup>33</sup> and Chu et al.,<sup>37</sup> who performed similar automatic segmentations for hip morphological parameter measurements. The proposed method achieved comparable or lower differences in measurement between manual and automatic workflows. Finally, comparison of the ICC with the manual interrater ICC by Harris-Hayes et al.<sup>50</sup> showed that the proposed method also performed comparable or better.

Figure 8 shows the Bland-Altman plots of all hip morphometry parameters. The femoral inclination and femoral version passed the Kolmogorov–Smirnov test for normality. In this case, the Limits of Agreement (LoA) were calculated as the mean +1.96 and –1.96 times the standard deviation. The alpha angle, acetabular version, acetabular coverage and LCE angles did not pass the test for normality and thus the 2.5th and 97.5th percentile were used to indicate the nonparametric LoA. These plots show that although the correlation

**TABLE 4** Comparison of the mean morphometric parameters found using the automatic and manual segmentations

Parameters	Manual segmentation		Automatic segmentation		Literature	
	Mean	SD	Mean	SD	Mean	SD
Alpha angle (°)	46.2	13.9	46.3	14.0	45.6 <sup>32</sup>	10.5 <sup>32</sup>
Femoral inclination (°)	129.5	5.2	129.7	5.4	129.2 <sup>32</sup>	6.2 <sup>32</sup>
Femoral version (°)	10.2	10.1	9.7	10.1	9.7 <sup>32</sup>	9.3 <sup>32</sup>
Acetabular version (°)	17.8	7.3	17.6	6.7	19 <sup>47</sup>	4 <sup>47</sup>
Acetabular coverage (%)	74.9	6.2	74.5	6.1	73 <sup>29</sup>	4 <sup>29</sup>
LCE angle (°)	32.0	8.5	31.3	8.2	33.6 <sup>48</sup>	18.1–48.0 <sup>a,48</sup>

Note: The interobserver variability is also shown to compare the range between observers and the manual and automatic segmentations.

<sup>a</sup>95% Confidence interval.

**TABLE 5** Differences and ICC(3,1) between hip morphometry parameters when calculated from manual and automatic segmentations

Parameters	Automatic and manual difference					Automatic and manual differences from literature		Manual interrater reliability ICC(2,1) <sup>49</sup>
	Mean	SD	Abs. Mean	Abs. SD	ICC	Abs. Mean	Abs. SD	
Alpha angle (°)	-0.27	8.1	2.5	7.2	0.822	-	-	0.78-0.86 <sup>50,a</sup>
Femoral inclination (°)	-0.3	1.2	0.72	0.95	0.986	2 <sup>33,a</sup> - 2.1 <sup>37</sup>	2 <sup>33,a</sup> - 1.6 <sup>37</sup>	0.96 <sup>50,a</sup>
Femoral version (°)	-0.1	0.9	0.55	0.71	0.998	1 <sup>33,a</sup> - 2.0 <sup>37</sup>	1 <sup>33,a</sup> - 1.5 <sup>37</sup>	0.97 <sup>50,a</sup>
Acetabular version (°)	0.02	1.8	0.95	1.5	0.983	-	-	0.94 <sup>50,a</sup>
Acetabular coverage (%)	0.4	1.5	0.94	1.1	0.986	2 <sup>33,a</sup> - 3.5 <sup>37</sup>	1 <sup>33,a</sup> - 2.3 <sup>37</sup>	-
LCE angle (°)	0.7	2.1	1.4	1.6	0.967	2 <sup>33,a</sup>	2 <sup>33,a</sup>	0.86 <sup>50,a</sup>

Note: The right side of the table shows the differences.

Abbreviations: CT, computerized tomography; MRI, magnetic resonance imaging.

<sup>a</sup>Based on MRI instead of CT segmentations.

between measurements was generally high, a few large outliers did occur, especially in the alpha angle measurements.

### 3.6 | NMDID dataset

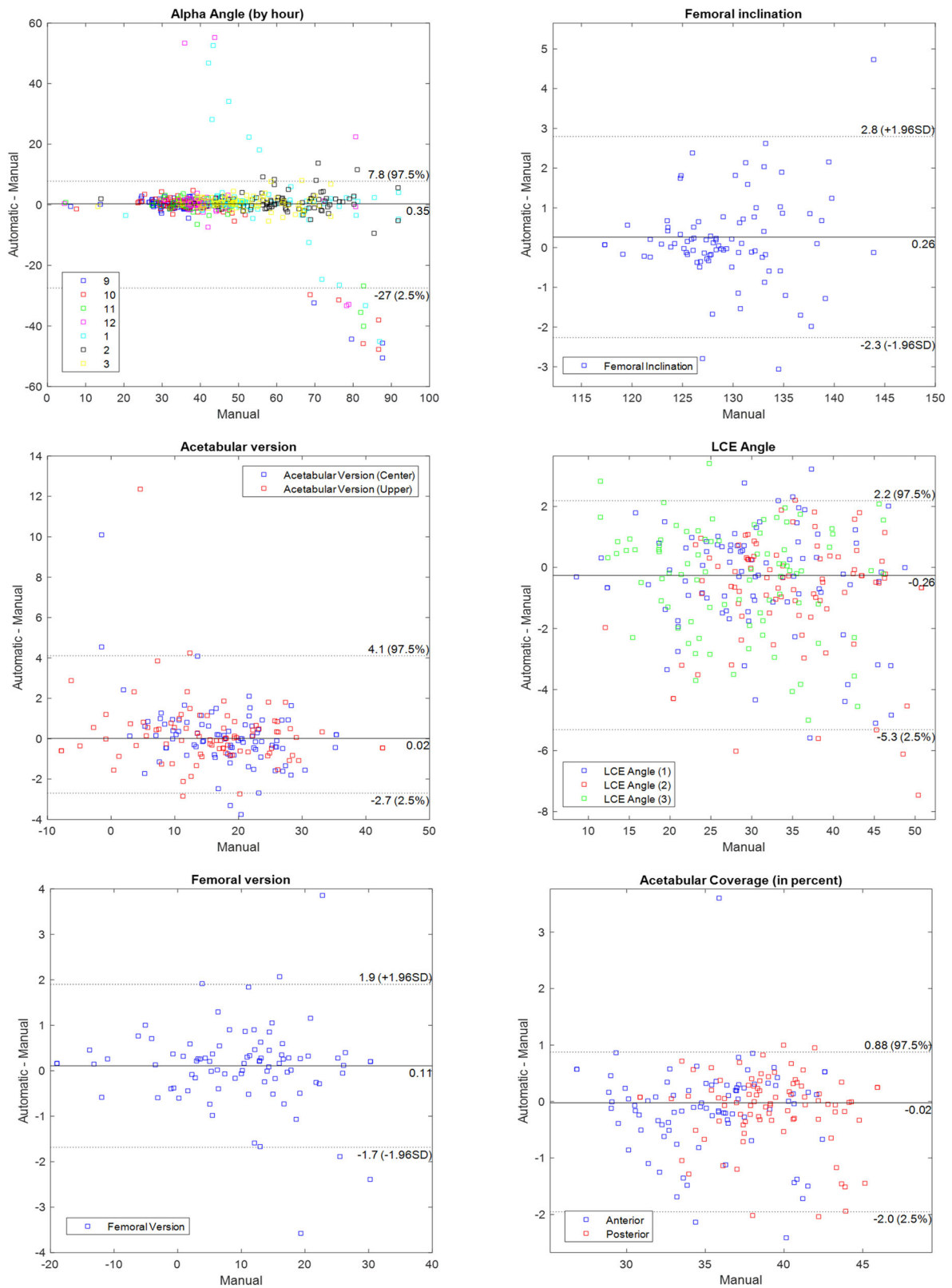
Results of the best performing network retrained on all 50 scans from the UMCU dataset and evaluated on the 10 scans of the NMDID dataset are shown in Table 3. The mean DSC was  $0.95 \pm 0.016$  and the HD95 was  $1.29 \pm 0.25$  mm.

## 4 | DISCUSSION

The aim of this study was to develop a fast, lightweight algorithm for semantic bone segmentation from CT images and assess its applicability to automatic 3D hip morphometry parameters. Segmentation of the bones from CT images has widespread use in the

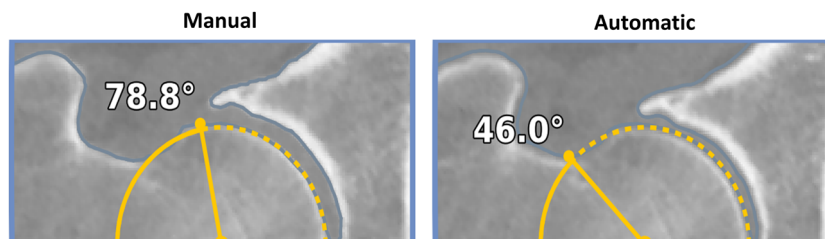
orthopaedic clinic, especially for the measurement of hip morphometry parameters. Therefore, many methods have been developed to segment the proximal femur and coxae.<sup>33,37-42,51</sup> However, only few studies have attempted to semantically segment larger parts of the human body with a single method,<sup>11,44</sup> which brings unique challenges such as computing time and memory constraints. In this study, we have shown that by using a cascaded network with a large receptive field we were able to achieve competitive semantic bone segmentation results while minimizing the time and memory requirements.

Our lightweight implementation of the V-net, with a lower number of convolutional layers, a smaller number of filters, and less down sampling layers and thus depth of the network, resulted in strongly reduced number of parameters. A single lightweight V-net without cascades had over 15 times less parameters than a similar V-net as described by Çiçek et al.<sup>17</sup> This allowed the network to run training and inference quickly and with relatively low memory consumption. The implementation of cascaded V-nets allowed the



**FIGURE 8** Bland-Altman plots showing the difference between automatic and manual segmentations for all hip morphometry measurements. The dotted lines indicated the Limits of Agreement, calculated as  $\pm 1.96$  times the standard deviation in case the values are normally distributed, or the 2.5th and 97.5th percentile otherwise

**FIGURE 9** Example of 12 o' clock alpha angle calculation that is highly influenced by a small change in segmentation



volume of the receptive field of the network to increase cubically with patch size, while only increasing the number of parameters and input voxels linearly.

The proposed method achieved a low MASD and HD on the UMCU dataset compared to other studies, with on average sub-voxel accuracy. The proposed method did perform worse on the HD metric for the coxae. Qualitative inspection of the segmentation of the coxae showed that the HD was especially influenced by segmentation errors on the border of the image, in scans where only part of the coxae were visible. A comparison to the cascaded nnU-net applied to our dataset showed that our network was able to attain better or comparable results with lower training and inference time, and lower GPU memory requirements.

Our method performed slightly worse on the NMDID dataset, but still outperformed the MASD results found in the literature on the tibia and fibula. The decreased accuracy can probably be attributed to significant differences between the UMCU and NMDID dataset. The subjects of the NMDID dataset were purposely chosen to be different in age than the UMCU patients, to test the robustness of the algorithm to different population demographics. Additionally, UMCU patients were alive while the NMDID dataset consists of recently deceased subjects, which may have introduced postmortem effects such as gas build-up in the lower abdomen.<sup>52</sup> The NMDID patients were also mostly scanned clothed, which introduced artefacts due to interference with objects on the body. Finally, the scanner and scanner settings differed. In future research, using a more diverse dataset for training might improve the robustness and thus general applicability of the network.

Excellent correspondence was found between the measurement of the HKA on the manual and automatic segmentations. A good or excellent correlation was also found for the hip morphometry parameters between the manual and automatic segmentations. Moreover, the correlation was higher than the manual interrater reliability reported by Harris-Hayes et al.<sup>50</sup> This is in accordance with the results of Zeng et al.,<sup>33</sup> who also used a deep learning method to automatically segment the hip and calculate the morphometry parameters. In contrast to their study, however, we segmented six bones instead of two, with a larger field of view that comprised the complete lower extremities. It should be noted that the studies by Zeng et al. and Harris-Hayes et al., with which our results were compared, were performed on MRI scans instead of CT scans. However, these represented the best results that could be found in the literature for comparison of automatic and manual hip morphometry assessment interrater reliability.

## 5 | LIMITATIONS

While the proposed methodology, just like nnU-net, is in principle generically applicable to other segmentation tasks, it has been evaluated on the specific challenge of bone segmentation in 3D CT, albeit with a large variety of osseous structures. Future work warrants more extensive evaluations to investigate the potential benefits of this approach to other segmentation challenges in different applications and on different data. In addition, this study focused on optimizing the number of U-nets in a cascaded approach with a fixed combination of other hyperparameters. An investigation of the optimal hyperparameters for each of the different network topologies might yield different results than presented here.

Although our segmentation method achieved good results relative to other state-of-the-art methods in the literature, care should be taken when comparing different studies. Each study reported on separate datasets with differing parameters such as inclusion criteria, scanner settings and voxel sizes.

We assumed that the morphological parameters that were calculated using the commercial software tool gave the correct results when presented with the manual segmentation. However, in some cases the manual segmentations included small spurious bone voxels on the femoral head, which increased the alpha angle more than would be clinically expected. An example of this is shown in Figure 9. These erroneous morphometry measurements were a limitation caused by our automatic approach, as these would normally be corrected manually when using the commercial software.

## 6 | CONCLUSION

In this study, we introduced a method for fully automatic lower extremity segmentation from CT. By using a cascaded V-net approach, it was possible to use information from a large receptive field, while maintaining a comparatively low computation time and GPU memory footprint. Comparison to state-of-the-art methods found in other studies showed that the segmentation accuracy also performed competitively. Furthermore, this study was the first to apply cascaded V-net based segmentation on lower extremity CT data, and validate the applicability to orthopaedic diagnosis using clinical implementations, i.e., HKA and hip joint measurements. Comparing the HKA and hip morphology between automatic and manual segmentations, the metrics showed good or excellent correlation, indicating that this method could be a valuable addition

to many orthopaedic applications that benefit from accurate bone segmentation.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the support of Zimmer Biomet with the support for implementation of the automatic hip morphological parameter calculation for this research. This work was partially funded by the Netherlands Organization for Scientific Research and is part of the Applied and Engineering Sciences research program, Project Number 15479. Peter Seevinck and Marijn van Stralen are minority shareholders at MRIguidance B.V. For the use of the data from the New Mexico Decedent Image Database, we acknowledge the Free Access Decedent Database funded by the National Institute of Justice Grant Number 2016-DN-BX-0144.

## AUTHOR CONTRIBUTIONS

**Ruurd Kuiper:** conceptualization; methodology; data curation; software; formal analysis; investigation; visualization; writing—original draft preparation. **Ralph Sakkers:** conceptualization; project administration; supervision; writing—review and editing. **Marijn van Stralen:** writing—review and editing. **Vahid Arbabi:** data curation; writing—review and editing. **Max Viergever:** project administration; supervision; writing—review and editing. **Harrie Weinans:** conceptualization; project administration; supervision; writing—review and editing. **Peter Seevinck:** conceptualization; project administration; supervision; writing—review and editing. All authors have read and approved the final submitted manuscript.

## ORCID

Ruurd J. A. Kuiper  <http://orcid.org/0000-0002-6511-3896>

Marijn van Stralen  <http://orcid.org/0000-0002-3051-5000>

Vahid Arbabi  <http://orcid.org/0000-0003-3347-2891>

Max A. Viergever  <http://orcid.org/0000-0003-2582-042X>

Harrie Weinans  <http://orcid.org/0000-0002-2275-6170>

## REFERENCES

- Lenchik L, Heacock L, Weaver AA, et al. Automated segmentation of tissues using CT and MRI: a systematic review. *Acad Radiol.* 2019; 26(12):1695-1706.
- Palit A, King R, Gu Y, et al. Prediction and visualisation of bony impingement for subject specific total hip arthroplasty. *Annu Int Conf IEEE Eng Med Biol Soc.* 2019:2127-2131.
- Lerch TD, Degonda C, Schmaranzer F, et al. Patient-specific 3-D magnetic resonance imaging-based dynamic simulation of hip impingement and range of motion can replace 3-D computed tomography-based simulation for patients with femoroacetabular impingement: implications for planning open hip preserv. *Am J Sports Med.* 2019;47(12):2966-2977.
- Fürnstahl P, Schweizer A, Graf M, et al. Computational Radiology for Orthopaedic Interventions. *Surgical treatment of long-bone deformities: 3D preoperative planning and patient-specific instrumentation.* Springer; 2016:123-149.
- Zhang YZ, Lu S, Zhang HQ, et al. Alignment of the lower extremity mechanical axis by computer-aided design and application in total knee arthroplasty. *Int J CARS.* 2016;11:1881-1890.
- Fürmetz J, Sass J, Ferreira T, et al. Three-dimensional assessment of lower limb alignment: accuracy and reliability. *Knee.* 2019;26(1): 185-193. Available from: [doi:10.1016/j.knee.2018.10.011](https://doi.org/10.1016/j.knee.2018.10.011)
- Renault JB, Aüllo-Rasser G, Donnez M, Parratte S, Chabrand P. Articular-surface-based automatic anatomical coordinate systems for the knee bones. *J Biomech.* 2018;80:171-178.
- Krishnan SP, Dawood A, Richards R, Henckel J, Hart AJ. A review of rapid prototyped surgical guides for patient-specific total knee replacement. *J Bone Jt Surg Ser.* 2012;B 94 B(11): 1457-1461.
- Minnema J, van Eijnatten M, Kouw W, Diblen F, Mendrik A, Wolff J. CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. *Comput Biol Med.* 2018;103: 130-139.
- van Eijnatten M, van Dijk R, Dobbe J, Streekstra G, Koivisto J, Wolff J. CT image segmentation methods for bone used in medical additive manufacturing. *Med Eng Phys.* 2018;51:6-16.
- Audenaert EA, Van Houcke J, Almeida DF, et al. Cascaded statistical shape model based segmentation of the full lower limb in CT. *Comput Methods Biomech Biomed Engin.* 2019;22(6):644-657.
- Asgari Taghanaki S, Abhishek K, Cohen JP, et al. Deep semantic segmentation of natural and medical images: a review. *Artificial Intelligence Review.* 2020;54:137-178. [doi:10.1007/s10462-020-09854-1](https://doi.org/10.1007/s10462-020-09854-1)
- Presciutti SM, Karukanda T, Lee M. Management decisions for adolescent idiopathic scoliosis significantly affect patient radiation exposure. *Spine J.* 2014;14(9):1984-1990.
- Lindgren Belal S, Sadik M, Kaboteh R, et al. 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.* 2019;113:89-95.
- Zhou Z, Zhao G, Kijowski R, Liu F. Deep convolutional neural network for segmentation of knee joint anatomy. *Magn Reson Med.* 2018;80(6):2759-2770.
- Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Lect. Notes Bioinformatics).* 2015; 9351:234-241.
- Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 3D U-net: learning dense volumetric segmentation from sparse annotation. 2016:424-432.
- Milletari F, Navab N, Ahmadi S-A 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016 :565-571 [cited 2020 Mar 16] Available from: <http://arxiv.org/abs/1606.04797>
- Noguchi S, Nishio M, Yakami M, Nakagomi K, Togashi K. Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques. *Comput Biol Med.* 2020; 121:103767.
- Berger L, Eoin H, Cardoso MJ, Ourselin S. An adaptive sampling scheme to efficiently train fully convolutional networks for semantic segmentation. *Commun Comput Inf Sci.* 2018;894:277-286.
- Kamnitsas K, Ledig C, Newcombe VFJ, et al. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med Image Anal.* 2017;36:61-78.
- Roth HR, Shen C, Oda H, et al. A multi-scale pyramid of 3D fully convolutional networks for abdominal multi-organ segmentation, 2018:417-425.
- Zhao H, Qi X, Shen X, et al. ICNet for Real-Time Semantic Segmentation on High-Resolution Images, 2018:418-434.
- Chen L-C, Papandreou G, Schroff F, Adam H 2017. Rethinking Atrous Convolution for Semantic Image Segmentation. Available from: <http://arxiv.org/abs/1706.05587>
- Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL. DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. *IEEE Trans Pattern Anal Mach Intell.* 2018;40(4):834-848.

26. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211.
27. Paley D. 2002. *Principles of Deformity Correction*. Berlin Heidelberg: Springer; Available from: <http://link.springer.com/10.1007/978-3-642-59373-4>
28. Jud L, Roth T, Frnstahl P, Vlachopoulos L, Sutter R, Fucentese SF. The impact of limb loading and the measurement modality (2D versus 3D) on the measurement of the limb loading dependent lower extremity parameters. *BMC Musculoskelet Disord*. 2020;21(1):418.
29. Dandachli W, Kannan V, Richards R, Shah Z, Hall-Craggs M, Witt J. Analysis of cover of the femoral head in normal and dysplastic hips: new CT-based technique. *J Bone Jt Surg Ser B*. 2008;90(11):1428-1434.
30. Khnlein W, Ganz R, Impellizzeri FM, Leunig M. Acetabular morphology: implications for Joint-preserving Surgery. *Clin Orthop Relat Res*. 2009;467(3):682-691.
31. Tannast M, Albers CE, Steppacher SD, Siebenrock KA 2011. Hip Pain in the Young Adult. *Eur. Instr. Lect*.
32. Toogood PA, Skalak A, Cooperman DR. Proximal femoral anatomy in the normal human population. *Clin Orthop Relat Res*. 2009;467(4):876-885.
33. Zeng G, Schmaranzer F, Degonda C, et al. MRI-based 3D models of the hip joint enables radiation-free computer-assisted planning of periacetabular osteotomy for treatment of hip dysplasia using deep learning for automatic segmentation. *Eur J Radiol Open*. 2021;8:100303.
34. Edgar H, Daneshvari Berry S, Moes E, et al. 2020. New Mexico Decedent Image Database. Available from: <http://nmdid.unm.edu/>
35. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15:1.
36. Wu D, Sofka M, Birkbeck N, Zhou SK. Segmentation of multiple knee bones from CT for orthopedic knee surgery planning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. 2014;17:372-380.
37. Chu C, Chen C, Liu L, Zheng G. FACTS: fully automatic CT segmentation of a hip joint. *Ann Biomed Eng*. 2015;43(5):1247-1259.
38. Kim JJ, Nam J, Jang IG. Fully automated segmentation of a hip joint using the patient-specific optimal thresholding and watershed algorithm. *Comput Methods Programs Biomed*. 2018;154:161-171.
39. Almeida DF, Ruben RB, Folgado J, et al. Fully automatic segmentation of femurs with medullary canal definition in high and in low resolution CT scans. *Med Eng Phys*. 2016;38(12):1474-1480.
40. Chu C, Bai J, Wu X, Zheng G. MASCG: multi-atlas segmentation constrained Graph method for accurate segmentation of hip CT images. *Med Image Anal*. 2015;26(1):173-184.
41. Chang Y, Yuan Y, Guo C, Wang Y, Cheng Y, Tamura S. Accurate pelvis and femur segmentation in hip CT with a novel patch-based refinement. *IEEE J Biomed Heal Informatics*. 2019;23(3):1192-1204.
42. Seim H, Kainmueller D, Heller M, et al. Automatic segmentation of the pelvic bones from CT data based on a statistical shape model. *Eurographics Workshop on Visual Computing for Biomedicine*. 2008;8:93-100.
43. Liu P, Han H, Du Y, et al. Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. *Int J Comput Assist Radiol Surg*. 2021;16(5):749-756.
44. Fu Y, Liu S, Li HH, Yang D. Automatic and hierarchical segmentation of the human skeleton in CT images. *Phys Med Biol*. 2017;62(7):2812-2833.
45. Mast NH, Impellizzeri F, Keller S, Leunig M. Reliability and agreement of measures used in radiographic evaluation of the adult hip. *Clin Orthop Relat Res*. 2011;469:188-199.
46. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420-428.
47. Hingsammer AM, Bixby S, Zurakowski D, Yen YM, Kim YJ. How do acetabular version and femoral head coverage change with skeletal maturity? *Clin Orthop Relat Res*. 2015;473(4):1224-1233.
48. Werner CM, Ramseier LE, Ruckstuhl T, et al. Normal values of Wiberg's lateral center-edge angle and Lequesne's acetabular index - a coxometric update. *Skeletal Radiol*. 2012;41(10):1273-1278.
49. Koo TK, Li MY. A Guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med*. 2016;15(2):155-163.
50. Harris-Hayes M, Commean PK, Patterson JD, Clohisy JC, Hillen TJ. Bony abnormalities of the hip joint: a new comprehensive, reliable and radiation-free measurement method using magnetic resonance imaging. *J Hip Preserv Surg*. 2014;1(2):62-70.
51. Avants BB, Tustison NJ, Wu J, Cook PA, Gee JC. ANTS: advanced open-source tools for normalization and neuroanatomy. *Neuroinformatics*. 2011;9(4):381-400. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21373993>
52. Wagensveld IM, Blokker BM, Wielopolski PA, et al. *Total-body CT and MR features of postmortem change in in-hospital deaths*. 2017. Available from: [doi:10.1371/journal.pone.0185115.t002](https://doi.org/10.1371/journal.pone.0185115.t002)
53. Kingma DP, Ba J 2015. Adam: A Method for Stochastic Optimization. International Conference for Learning Representations. 1-15; Available from: <http://arxiv.org/abs/1412.6980>.

## SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

**How to cite this article:** Kuiper RJA, Sakkers RJB, Stralen M, et al. Efficient cascaded V-net optimization for lower extremity CT segmentation validated using bone morphology assessment. *J Orthop Res*. 2022;40:2894-2907. [doi:10.1002/jor.25314](https://doi.org/10.1002/jor.25314)