BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 69(3), 2021, Article number: e136751
DOI: 10.24425/bpasts.2021.136751

Segmentation of bone structures with the use of deep learning techniques

Zuzanna KRAWCZYK* and Jacek STARZYŃSKI

Warsaw University of Technology, ul. Koszykowa 75, 00-662 Warsaw, Poland

Abstract. The paper is focused on automatic segmentation task of bone structures out of CT data series of pelvic region. The authors trained and compared four different models of deep neural networks (FCN, PSPNet, U-net and Segnet) to perform the segmentation task of three following classes: background, patient outline and bones. The mean and class-wise Intersection over Union (IoU), Dice coefficient and pixel accuracy measures were evaluated for each network outcome. In the initial phase all of the networks were trained for 10 epochs. The most exact segmentation results were obtained with the use of U-net model, with mean IoU value equal to 93.2%. The results where further outperformed with the U-net model modification with ResNet50 model used as the encoder, trained by 30 epochs, which obtained following result: mIoU measure – 96.92%, "bone" class IoU – 92.87%, mDice coefficient – 98.41%, mDice coefficient for "bone" – 96.31%, mAccuracy – 99.85% and Accuracy for "bone" class – 99.92%.

Key words: deep learning; semantic segmentation; U-net; FCN; ResNet; computed tomography.

1. Introduction

Individualised bone models play an important role in various medical applications such as orthopaedic operation planning [1, 2], motion visualisation [3] or radiotherapy treatment planning [4]. The traditional way of bone model creation is a manual or semi-manual time-consuming process. One of its key steps is the precise segmentation of bone structures out of a series of medical images.

In recent years the use of deep convolutional neural networks is becoming more popular in the field of computer vision and image processing as well as in medical applications, including segmentation of anatomical structures out of CT data [5, 6].

The aim of this paper is the verification of usefulness of popular deep model networks, used for general purpose segmentation to the task of bone structure segmentation out of CT data series. The automatic deep learning based bone segmentation can be further used in the method of automatic femur bone creation described in [7], replacing its current bone edge detection phase, based on state of the art image processing algorithms. Four models of neural networks are compared: FCN [8], PSPNet [9], U-net [10] and Segnet [11]. During the segmentation phase network assigns each pixel in the image to the 3 following classes: background, patient outline and bones.

The article is organized as follows: the second chapter discusses the current state of knowledge on segmentation of medical images, the third chapter briefly characterizes four models of neural networks that are used in the work. The fourth chapter describes the set of data used to train the networks. The next chapter presents the results obtained. The last chapter focuses on conclusions and perspectives for further work.

The preliminary results of the work were presented in the CPEE 2020 conference [12]. The paper broadens the subject with segmentation performed by FCN network and modifications of the U-net network. Evaluation of the results was performed on much larger dataset with the use of additional fitting measures.

2. Literature review and goal

Recent publications reported that deep neural networks are capable to perform segmentation task on medical images with the accuracy comparable to human expert. One of the common tasks performed successfully by the deep neural network is organ segmentation. In [13] novel superpixel-based and boundary sensitive CNN was used for liver segmentation out of CT data and achieved high Dice coefficient value equal to $97.31 \pm 0.36\%$. Authors of [14] successfully performed segmentation of head and neck organs at risks utilising 3D U-net architecture with performance similar to the expert.

The segmentation process is also used in detection of pathological changes in the human body. Authors of [15] performed segmentation of polycystic kidneys with mean volume difference $0.68 \pm 2.2.\%$ in comparison to expert segmentation. Various deep convolutional neural networks are also utilised to the task of automatic brain tumour segmentation [16].

In addition to the CT and MRI imaging analysis, deep neural networks are also used in the processing of histopathological images for example in segmenting and classifying epithelial and stromal regions [17] or in damaged tissue detection [18].

Several papers tackled the subject of bone structure segmentation. In [19] a custom deep learning method was developed for segmentation of 49 bones out of PET/CT datasets.

*e-mail: zuzanna.krawczyk@ee.pw.edu.pl

Manuscript submitted 2020-10-20, revised 2020-12-22, initially accepted for publication 2021-01-06, published in June 2021

The training set consisted of 100 data series. In [20] the suitability of the U-net network to the bone segmentation task was shown obtaining a Dice coefficient value of 92% and an IoU value equal to 85%. The method for creation of STL skull models described in [21] is also based on convolutional neural network.

Alternative approaches to segmentation of biological structures do not use CNNs, however they do also often incorporate additional knowledge about the shape topology of segmented structures. For example, in [22] reconstruction of overlapping cells in breast cancer FISH images with the use of PatchMatch algorithm and database of cell shapes was performed. Another popular approach is Statistical Shape Model used in [23] for the segmentation of pelvic and femur bones.

Encouraged by the reports about the effectiveness of deep learning networks the authors aim to use CNN as an alternate and possibly better approach to creation of 3D skeleton models out of CT data. To check this possibility we evaluated four different models of neural networks to segment pelvis region.

3. Models used

The section briefly describes neural network models used in the paper.

3.1. FCN. Fully convolutional network (FCN), described in [8] consists only of convolutional layers (it does not contain any fully connected layers). The network can be trained to create the segmentation of an image of any size. In order to obtain the output segmentation image of the same size as the input one, after the series of convolutions and pooling layers the image is upsampled by using the deconvolutional layers.

Two types of FCN networks are considered in this paper: FCN32 and FCN8 network. In FCN32 network, after being processed by several convolution layers, the features map is upsampled 32 times to obtain the output size of the image. The output of the network may be rough because of the loss of the information about the spacial location, during pooling operation. To prevent this FCN8 network upsamples and combines the feature maps of higher resolution, from previous layers, which gives information on both general and specific features of the image.

3.2. PSPNet. Pyramid Scene Parsing Network [9] (PSPNet) introduces pyramid pooling module which allows to incorporate global context information into the network knowledge. It also uses dilate convolutional layers.

3.3. U-net. U-Net [11] is FCN based segmentation network with symmetric encoder and decoder parts which give the network architecture a characteristic U-like shape. The U-net network also provides the skip connections between corresponding parts of encoder and decoder, using the concatenation operator instead of sum (like in case of FCN network). In combination with extensive data augmentation U-net network showed to give good results in segmentation of biomedical images.

3.4. Segnet. Segnet [11] network uses similar encoder-decoder architecture as the FCN network, but introduces more skip connections than FCN. In the encoder (downsampling) part of the network the indices of max-pooling operation are stored and further used in decoder part, during the upsample operation instead of copying features. This makes Segnet network more memory efficient than FCN and U-net networks.

4. Dataset used

The training dataset consists of 987 images extracted from 10 CT series of pelvis and its surrounding. 97 images from the set were considered as validation data. 730 images taken from 5 different CT series, separate from training data, were used as a testing data. Example slices from the training data are presented in Fig. 1. Corresponding labels map was created for each of the training images classifying each pixel to one of 3 classes: background (class 0), patient outline (class 1) and bone (class 2).

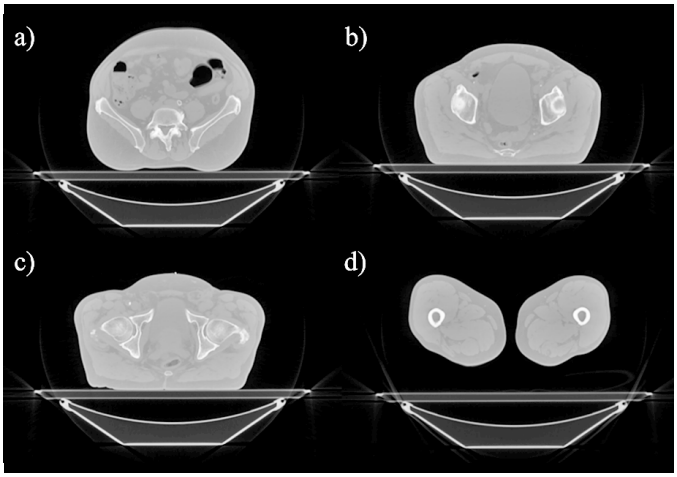

Fig. 1. Example CT slices from training data presenting cross-section of a) sacrum and ilium b) sacrum and ilium c) ischium and upper extremity of femur d) body of femur

The number of images was additionally augmented by geometrical operations like scaling, translation or rotation as well as by changing image contrast, performing histogram equalisation or gaussian blur. Examples of augmented data are shown in Fig. 2.

The heavy augmentation procedure from [24] based on operations implemented in [25] was applied to the set of training images. The following operations were applied most frequently: vertical flip – in 20% of images, horizontal flip, cropping and padding by –5% to 10% (50% of images), the sequence of scaling (from 80% to 120%), rotation (–45° to 45° degrees) and shear (from –16° to 16° degrees) – in 50% of images. 0 to 5 following operations were randomly performed per each image: blurring operation (gaussian (σ parameter with value from 0 to 3), average or median (kernel size form 2 to 7) blur-

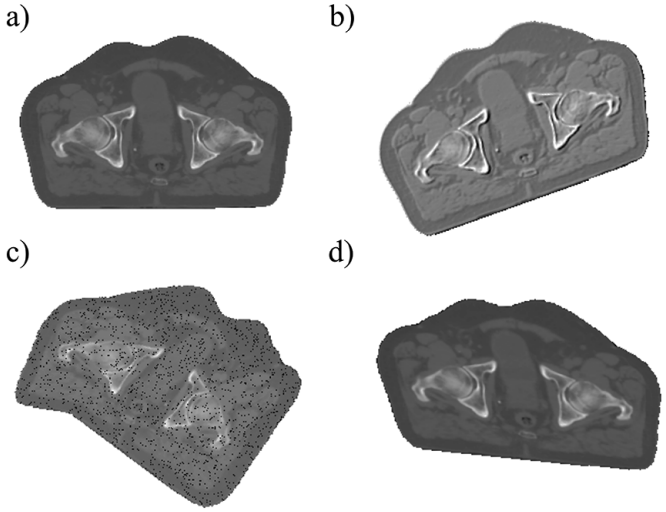

Fig. 2. Examples of data augmentation: a) original data, data after: b) rotation and contrast change c) salt, and non-affine deformation, d) rotation and scaling. Pixels belonging to "background" class were set to the white colour to increase readability of the image

ring), image sharpening, edge detection and addition of them to the image, additive gaussian noise, removal of up to 10% of image pixels, inversion of the pixel values (in 5% of images), change of the brightness of the image by –10% to 10% of the original value, moving pixels locally with different strength, image distortion by applying the affine geometric transformations to the local neighbourhoods of pixels, the random four-point perspective transformation of the image. The sequence of above operations is applied in random order to the image set, 10 times to each image.

All CT images have size 512×512 pixels and pixel spacing equal to 1.5625 mm. The distance between neighbouring images in the series (slices) was equal to 2.5 mm.

The data were obtained due tu courtesy of Maria Sklodowska-Curie National Research Institute of Oncology.

5. Results

5.1. Network training. In order to verify the effectiveness of deep neural networks for the task of bone segmentation out of CT data 4 different deep neural network architectures were tested: FCN, PSPNet, U-net and Segnet. We trained network models available in the library [24].

The code was written with the use of Keras library (version 2.3.1) [26] with Tensorflow (version 1.5) [27] backend and launched in Jupyter notebook environment. In order to speed up the training step it was performed on GPU (NVIDIA GeForce RTX 2080 Ti).

All of the networks were trained on the data set described in Section 4 for 10 epochs with 512 steps per epoch. The loss function was categorical cross-entropy. As optimiser the default stochastic gradient descent method which implements "adadelta" algorithm, was used.

5.2. Segmentation evaluation measures. During the training step, the value of pixel accuracy (1) was registered:

$$accuracy = \frac{TP}{TP + TN + FP + FN}, \tag{1}$$

where accuracy measure can be defined as the number of pixels correctly predicted by the network divided by the number of all pixels. TP (true positive) is the number of pixels correctly assigned to a given class; TN (true negative) is a number of pixels which are correctly predicted as not belonging to a given class. FP (false positive) denotes pixels which were wrongly classified as belonging to given class and FN (false negative) is the number of pixels which network missed to assign to the class.

Example plot of overall accuracy measure for U-net network is depicted in Fig. 3.

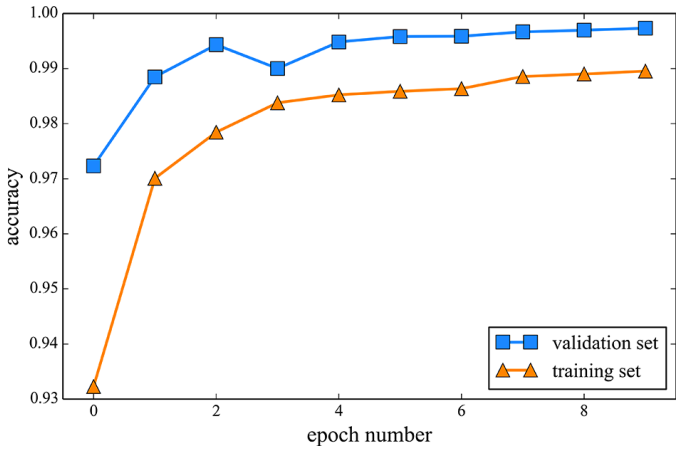

Fig. 3. U-net: accuracy value of the validation (blue colour) and training set (orange colour) for subsequent epochs

The value of mean Intersection over Union (IoU) measure as well as per class IoU value were evaluated for the validation set, for each training epoch. IoU measure is understood as the number of common pixels from given class, between the reference image (P_R) and predicted image (P_P) divided by the union of the pixels from reference and predicted image.

$$IoU = \frac{P_R \bigcap P_P}{P_R \bigcup P_P} = \frac{TP}{TP + FN + FP}. \tag{2}$$

The example of IoU value for consecutive training epochs, for U-net model is presented in Fig. 4. The IoU value for "background" and "patient outline" class achieves value greater than 90% in early phase of the training (after the first epoch for "background" class and after the third epoch for "patient outline" class), and does not change rapidly in the following epochs. The IoU value for the "bone" class changes dynamically until the 5th epoch and then it flattens on the value equal

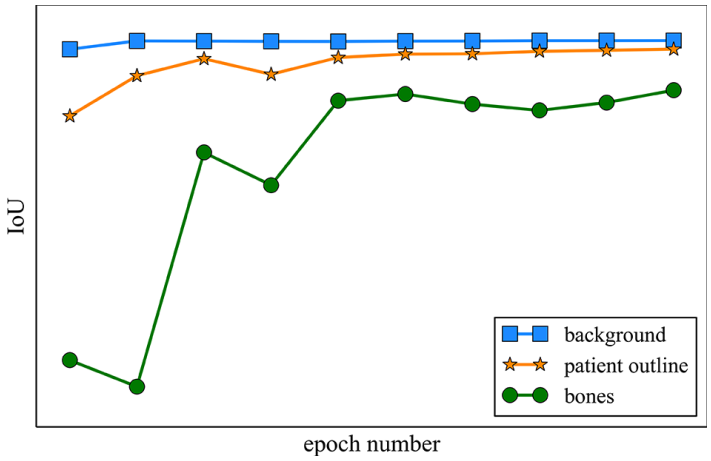

Fig. 4. U-net: IoU value of the validation set for subsequent epochs for 3 classes: background (blue colour), patient outline (orange colour), bones (green colour)

to approximately 80%. The lower value of IoU measure can be caused by the small representation of pixels belonging to the "bone" class in comparison to the first two classes.

Third measure used in further evaluation of network models is Dice coefficient (F1 score) (3) which is equal to twice the number of common pixels from a given class in the reference image and predicted image divided by the sum of the pixels belonging to the class from reference and predicted image:

$$Dice_{coeff} = \frac{2|P_R \bigcap P_P|}{|P_R| + |P_P|} = \frac{2TP}{2TP + FN + FP}. \qquad (3)$$

The measure is positively correlated with the IoU measure, however in IoU measure case individual bad segmentation results influence the overall evaluation of the model more severely (Fig. 5).

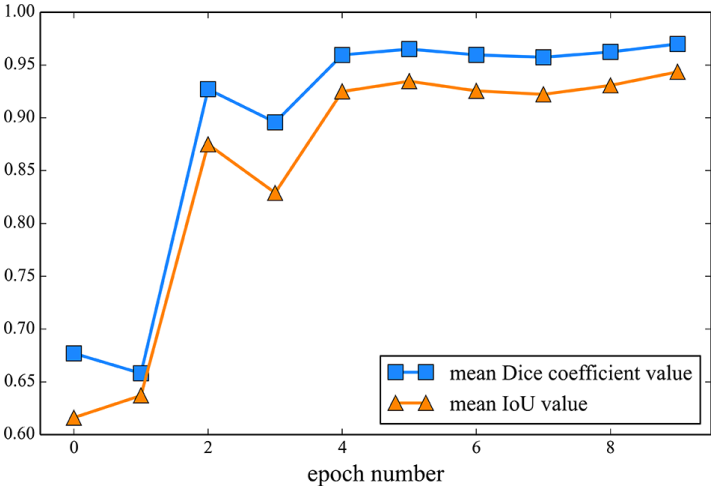

Fig. 5. U-net: mean IoU value (orange colour) and mean Dice coefficient value (blue colour) of the validation set for subsequent epochs

5.3. Evaluation of the segmentation results. As IoU measure is considered as the most strict, the network weights established for the epoch with the highest mean IoU, where chosen for further computations. Then, for weights chosen in such a way the IoU, Dice coefficient and pixel accuracy measures were computed for separate test data set. The detailed performance of each network is summarised in the Table 1.

Table 1
IoU, Dice coefficient and Pixel accuracy measures evaluated for segmentation results obtained by FCN, PSPNet, U-net and Segnet networks

Name	It. No.	Measure [%]	Mean value	Classwise measure		
				0	1	2
FCN8	10	IoU	92.39	99.78	96.45	80.95
		Dice Coeff.	95.85	99.89	98.19	89.47
		Pixel Acc.	99.72	99.81	99.58	99.78
FCN32	10	IoU	82.83	99.29	90.12	59.09
		Dice Coeff.	89.58	99.64	94.80	74.28
		Pixel Acc.	99.22	99.37	98.83	99.46
PSPNet	6	IoU	83.26	99.61	92.78	57.38
		Dice Coeff.	89.66	99.81	96.26	72.92
		Pixel Acc.	99.42	99.66	99.14	99.48
U-net	10	IoU	93.91	99.76	96.65	85.32
		Dice Coeff.	96.75	99.88	98.30	92.08
		Pixel Acc.	99.74	99.79	99.61	99.83
Segnet	7	IoU	90.30	99.67	95.30	75.93
		Dice Coeff.	94.58	99.84	97.60	86.32
		Pixel Acc.	99.63	99.71	99.45	99.73

The example segmentation performed by each of the models is depicted in Fig. 6.

All network models obtained mean IoU value higher than 83%, and mean Dice coefficient value higher than 89% but only 3 models resulted with the IoU value over 90% – FCN8, U-net and Segnet. Background and patient classes were segmented in all cases with high accuracy. The greatest discrepancies are noticeable in the case of, most important, "bone" class. The IoU values for "bone" class vary from 57.38% to 85.32%, whereas Dice coefficient ranges from 72.92% to 92.08% for PSPnet and U-net models respectively. Non-intuitively, the lowest IoU value was obtained by PSPNet which is the only tested network incorporating global context into the segmentation task. The pixel accuracy measure in all cases resulted with the value grater than 99% – which is not reliable result, because of the imbalanced ratio of true negative to true positive pixel detections. (Especially in case of "bone" class).

Example segmentation of upper part of the pelvis bone presented in Fig. 6 shows that only FCN8 model and U-net network segmentation results have shape similar to the ref-

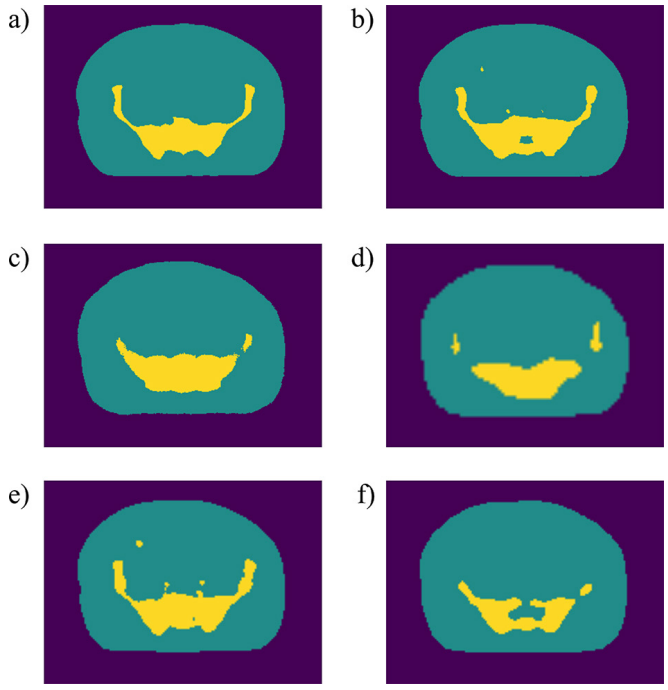

Fig. 6. Examples of 3-class segmentation results: a) reference segmentation, b) FCN8 model, ”bone” class IoU: 73.72% , c) FCN32 model, ”bone” class IoU: 63.66%, d) PSPNet model, ”bone” class IoU: 58.74% e) U-net model, ”bone” class IoU: 79.75%, Segnet model, ”bone” class IoU: 60%

erence one. In other cases pelvis wings are not segmented or under-segmented.

The most memory efficient model (with the smallest number of parameters) is PSPNet followed by the Segnet and U-net networks. The mean time of prediction of one 512×512 pixels image is similar for those three models and vary from 23.3 ms for Segnet to 26.3 ms for U-net model. Both FCN networks have an order of magnitude larger number of parameters and thus are much more memory expensive, they also need significantly more time to perform the segmentation. The detailed number of parameters and mean execution time for each of the discussed models is presented in Table 2.

Table 2
Models mean execution time and parameters number

Model	Mean execution time [ms]	No. of parameters	No. of trainable parameters
FCN8	57.6	69,736,873	69,734,953
FCN32	59.5	69,769,603	69,767,683
PSPNet	23.3	3,282,819	3,275,779
U-net	26.3	4,472,323	4,468,995
SegNet	24.4	3,698,179	3,694,851

The IoU value obtained for “bone” class by U-net model is over 4% higher than the result obtained by the model second in line (FCN8) and over 9% higher than SegNet model, which obtained third best mean IoU score. Moreover the model is relatively compact and fast in prediction process. Thus, the U-net network model is considered best suited for bone structures segmentation task.

For a single test image the lowest IoU value for “bone” class obtained by U-net was equal to 57.49% and the highest one was equal to 98.39 %, with mean IoU value equal to 87.15% and standard deviation 5.98%.

Example segmentations performed by U-net model for various test images are shown in Fig. 7. The most exact segmentation results are obtained for diaphysis part of femur bone visible in the first row of the image (“bone” class IoU value equal to 98.38%). In further examples some regions are slightly over or under-segmented. In case of the image from the third row of Fig. 7 markers used in oncological treatment were wrongly classified as bone structure. Small separated groups of pixels were also incorrectly classified as bone in image from row number 4.

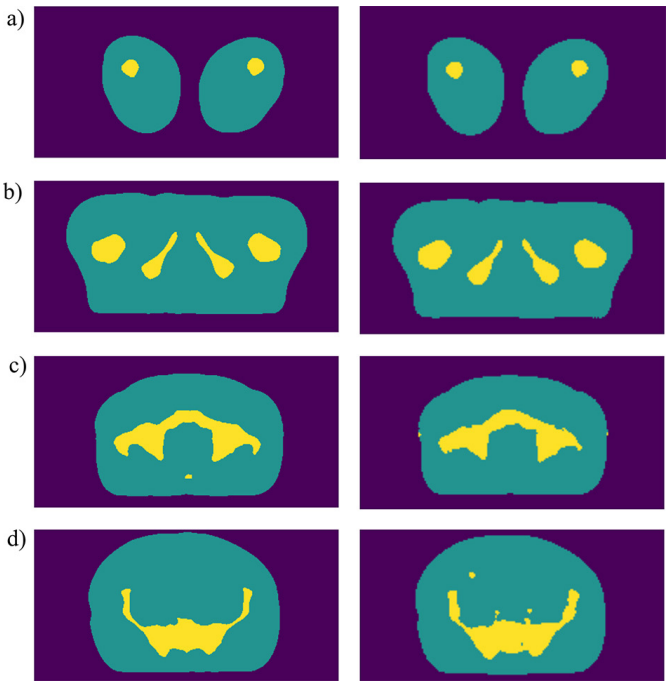

Fig. 7. Examples of 3-class segmentation results: reference segmentation – left column; segmentation performed by U-net model – right column. a) Mean IoU: 98.69%, ”bone” class IoU: 98.38%, b) Mean IoU: 95.81%, ”bone” class IoU: 91.63%, c) Mean IoU: 93.5%, ”bone” class IoU: 85.2%, d) Mean IoU: 93.5%, ”bone” class IoU: 79.75%

5.4. U-net modification. In the further step two modifications from [24] of the classical U-net model are considered in order to improve the accuracy of the results. The encoder of the U-net network is replaced by the VGG [29] model and Resnet [28] model. In both cases models pre-trained on ImageNet [30] database were applied. The weights used in VGG and Resnet models can be found in [31] and [32] respectively.

VGG network was created by Visual Geometry Group from University of Oxford. VGG is a CNN model based on Alex-

Net which uses set of small receptive fields (of size 3×3 with the stride equal to 1) instead of large one as well as multiple non-linear ReLUs, which allowed to create deeper network, with ability to learn more complex features. VGG has also less parameters than AlexNet.

ResNet (Residual Network) was the first architecture which allowed to efficiently train very deep neural networks (even with over 150 layers) by introducing the skip connection concept. The Resnet50 version of the network was used.

The networks were trained for 30 epochs and the weights from epoch with the best IoU score computed for validation set were chosen for further evaluation. The standard U-net network was trained for 30 epochs as well, to allow the further comparison. The detailed comparison of the models is presented in Table 3.

Table 3
IoU, Dice coefficient and Pixel accuracy measures evaluated for segmentation results obtained by VGG-U-net and Resnet-Unet

Encoder	It. No.	Measure [%]	Mean value	Classwise measure		
				0	1	2
U-net	30	IoU	94.64	99.83	97.41	86.67
		Dice Coeff.	97.16	99.92	98.69	92.86
		Pixel Acc.	99.80	99.85	99.70	99.84
VGG	28	IoU	96.55	99.85	98.05	91.76
		Dice Coeff.	98.21	99.92	99.02	95.70
		Pixel Acc.	99.85	99.87	99.78	99.91
Resnet	20	IoU	96.92	99.83	98.05	92.87
		Dice Coeff.	98.41	99.91	99.01	96.31
		Pixel Acc.	99.85	99.85	99.77	99.92

The longer training did not improve the results obtained by the standard U-net network significantly resulting with 0.73% better mean IoU value and by 1.35% greater IoU for "bone" class. Both networks using pre-trained backbones significantly outperformed traditional U-net model with over 90% IoU measure for "bone" class. The best result was achieved by the Resnet-U-net model giving 92.87% IoU accuracy for the class of interest.

Table 4
IoU measure calculated for each CT dataset separately, for Resnet-U-net model

CT set No.	Img. No.	mIoU [%]	Classwise measure		
			0	1	2
1	85	97.10	99.88	98.59	92.83
2	152	97.10	99.87	98.32	93.10
3	124	96.25	99.90	97.96	90.87
4	143	96.79	99.88	98.27	92.22
5	226	97.04	99.71	97.66	93.75
Combined	730	96.92	99.83	98.05	92.87

The IoU measure was additionally calculated for each CT data serie used in the test data, separately, for the best performing network, in order to find out if different examples are segmented with similar accuracy. The detailed results are presented in Table 4. The values of mean IoU measure obtained for all 5 sets differ less than 0.86% and in case of "bone class" the difference between the worst and best segmentation case is equal to 2.88%. The bone structures in all sets were segmented with over 90% IoU precision which can indicate that model is capable to properly perform segmentation for various input data.

6. Conclusion and future work

The paper compared 4 models of deep neural networks applied to the task of bone segmentation out of CT data. The most exact results were obtained by U-net network with 93.91% mean IoU value and 85.32% IoU value for "bone" class. The network is also more compact and faster than FCN8 model which obtained second best prediction result.

The U-net network analysis was enhanced by replacing its standard encoder with two different backbones – VGG and Resnet50 models. Both networks outperformed standard U-net model with Resnet50 backbone yelding the best result with mIoU measure – 96.92%, "bone" class IoU – 92.87%, mDice coefficient – 98.41% and Dice coefficient for "bone" class equal to 96.31%.

Several, recent publications tackled the subject of various bone segmentation using deep convolutional neural network. The authors of [20] performed bone segmentation based on U-net inspired network as well. The network was trained with 8 whole-body CT scans (6800 axial slices). The obtained mIoU value was equal to 91%, the mDice coefficient value was 95%, which are lower values than in case of U-net-Resnet model. afiga141!@tion, much wider range of anatomical regions was segmented. In case of pelvic and femur regions the network described in [15] obtained better than average results – presented only in bar diagram form – the value of mDice coefficient approximately corresponds to the one obtained by our U-net-Resnet model.

The CNN network presented in [21], trained on 20 CT series performed segmentation of skull with mean Dice similarity coefficient of 92% between the prediction and golden standard.

In [33] the U-net network prediction combined with preprocessing techniques resulted in segmentation of bone structures in whole body CT scans with the value of mDice coefficient equal to 97.9% and 96.5% for two in-house data-sets and 93.4% for external dataset.

In [19] results of segmentation of 49 bones out of CT data series are presented. The authors implemented three-step procedure which combines the use of CNN and active shape method. The training set consisted of 100 CT series. The results were compared with manual segmentation of 5 different bones in 5 test CT series. The values of Dice coefficient obtained for each type of bone were: 86%(Th 7), 85% (L3), 88% (sacrum), 84% (7th rib) and 83%(sternium).

The Dice coefficient value between golden standard and segmentation performed by neural networks in discussed works varies from 83% to 97.9%, where only predictions obtained in [33] resulted in higher value of mDice coefficient then described U-net-Resnet model. The obtained results cannot be directly compared because they apply to different range of bone structures of various anatomical characteristics. In case of biological structures even segmentations performed by human experts differ from one another, thus the similarity measure between two segmentation results will almost never be equal to 100%. Values of mDice coefficient obtained by U-net-Resnet segmentation, similar to ones obtained in different papers indicate that segmentation result described in our paper is on a satisfactory level.

In further research the authors would like to incorporate the U-net-Resnet based segmentation step into the automatic bones model creation framework [7]. The prior step will include enriching the training data set by providing additional images of the pelvis bone and upper part of the femur bone – thus they are under-represented and more anatomically complex in comparison to the images of the lower part of the femur.

The next step will include constraining the segmentation space to the bone area by performing first the bone detection by YOLO network.

Finally, the network can be trained to recognise specific bone structures such as femur or plevis.

Nevertheless, obtained results achieved good accuracy and are promising.

Acknowledgment. The authors would like to thank Maria Sklodowska-Curie National Research Institute of Oncology for the medical data used in research.

References

[1] E. Stindel, *et al.*, "Bone morphing: 3D morphological data for total knee arthroplasty" *Comput. Aided Surg*. 7(3), 156–168 (2002), doi: 10.1002/igs.10042.

[2] F. Azimifar, K. Hassani, A.H. Saveh, and F.T. Ghomsheh, "A medium invasiveness multi-level patient's specific template for pedicle screw placement in the scoliosis surgery", *Biomed. Eng. Online* 16, 130 (2017), doi: 10.1186/s12938-017-0421-0.

[3] L. Yahia-Cherif, B. Gilles, T. Molet, and N. Magnenat-Thalmann, "Motion capture and visualization of the hip joint with dynamic MRI and optical systems", *Comp. Anim. Virtual Worlds* 15, 377–385 (2004).

[4] V. Pekar, T.R. McNutt, and M.R. Kaus, "Automated modelbased organ delineation for radiotherapy planning in prostatic region", *Int. J. Radiat. Oncol. Biol. Phys*. 60(3), 973–980 (2004).

[5] D. Ravì, *et al.*, "Deep learning for health informatics," *IEEE J. Biomed. Health. Inf*. 21(1), 4–21 (2017), doi: 10.1109/JBHI.2016.2636665.

[6] G. Litjens, *et al.*, "A survey on deep learning in medical image analysis", Med. Image Anal. 42, 60–88 (2017), doi: 10.1016/j.media.2017.07.005.

[7] Z. Krawczyk and J. Starzyński, "YOLO and morphingbased method for 3D individualised bone model creation", *2020 International Joint Conference on Neural Networks (IJCNN)*, Glasgow, United Kingdom (2020), doi: 10.1109/IJCNN48605.2020.9206783.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, 3431–3440 (2015), doi: 10.1109/CVPR.2015.7298965.

[9] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, 6230–6239 (2017), doi: 10.1109/CVPR.2017.660.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation", in *Navab N., Hornegger J., Wells W., Frangi A. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. MICCAI 2015. Lecture Notes in Computer Science*, vol. 9351, Springer, Cham. (2015), doi: 10.1007/978-3-319-24574-4_28.

[11] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation", *IEEE Trans. Pattern Anal. Mach. Intell*. 39(12), 2481–2495 (2017), doi: 10.1109/TPAMI.2016.2644615.

[12] Z. Krawczyk and J. Starzyński, "Deep learning approach for bone model creation", *2020 IEEE 21st International Conference on Computational Problems of Electrical Engineering (CPEE)*, (2020), doi: 10.1109/CPEE50798.2020.9238678.

[13] W. Qin, J. Wu, F. Han, Y. Yuan, W. Zhao, B. Ibragimov, J. Gu, and L. Xing, "Superpixel-based and boundary-sensitive convolutional neural network for automated liver segmentation", *Phys. Med. Biol*. 63(9), 95017 (2018), doi: 10.1088/1361–6560/aabd19.

[14] S. Nikolov, *et al.*, "Deep learning to achieve clinically applicable segmentation of head and neck anatomy for radiotherapy", Technical Report, *ArXiv*, (2018), doi: arXiv:1809.04430.

[15] T.L. Kline, *et al.*, "Performance of an artificial multi-observer deep neural network for fully automated segmentation of polycystic kidneys", *J Digit Imaging* 30, 442–448 (2017), doi: 10.1007/s10278-017-9978-1.

[16] A. Wadhwa, A. Bhardwaj, and V.S. Verma, "A review on brain tumor segmentation of MRI images", Magn. Reson. Imaging 61, 247–259 (2019), doi: 10.1016/j.mri.2019.05.043.

[17] J. Xu, X. Luo, G. Wang, H. Gilmore, and A. Madabhushi, "A Deep Convolutional Neural Network for segmenting and classifying epithelial and stromal regions in histopathological images", Neurocomputing 191, 214–223 (2016), doi: 10.1016/j.neucom.2016.01.034.

[18] Z. Swiderska-Chadaj, T. Markiewicz, J. Gallego, G. Bueno, B. Grala, and M. Lorent, "Deep learning for damaged tissue detection and segmentationin Ki-67 brain tumor specimens based on the U-net model", *Bull. Pol. Acad. Sci. Tech. Sci.* 66(6), 849–856 (2018), doi: 10.24425/bpas.2018.125932.

[19] S. Lindgren Belal, 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*. 113, 89–95 (2019), doi: 10.1016/j.ejrad.2019.01.028.

[20] A. Klein, J. Warszawski, J. Hillengaß, and K.H. Maier-Hein, "Automatic bone segmentation in whole-body CT images", *Int J Comput Assist Radiol Surg*. 14(1), 21–29 (2019), doi: 10.1007/s11548-018-1883-7.

[21] J. Minnema, M. van Eijnatten, W. Kouw, F. Diblen, A. Mendrik, and J. Wolff, "CT image segmentation of bone for medical additive manufacturing using a convolutional neural network", *Comput. Biol. Med*. 103, 130–139 (2018), https://doi.org/10.1016/j.compbiomed.2018.10.012.

[22] T. Les, T. Markiewicz, T. Osowski, and M. Jesiotr, "Automatic reconstruction of overlapped cells in breast cancer FISH images", *Expert Syst. Appl*. 137, 335–342 (2019).

[23] F. Yokota, T. Okada, M. Takao, N. Sugano, Y. Tada, and Y. Sato, "Automated segmentation of the femur and pelvis from 3D CT data of diseased hip using hierarchical statistical shape model of joint structure", *Med Image Comput Comput Assist Interv.*, 811–818 (2019), doi: 10.1007/978-3-642-04271-3_98.

[24] D. Gupta, "Semantic segmentation library", accessed 19-Jan-202, [Online], Available: https: //divamgupta.com/image-segmentation/2019/06/06/ deep-learning-semantic-segmentation-keras.html.

[25] A.B. Jung, *et al.*, "Imgaug library", accessed 01-Feb-2020, [Online], Available: https://github.com/aleju/imgaug (2020).

[26] F. Chollet, *et al.*, "Keras", [Online], Available: https://keras.io, (2015).

[27] M. Abadi, *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous systems", [Online], Available: tensorflow.org, (2015).

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 770–778 (2016), doi: 10.1109/CVPR.2016.90.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *CoRR*, (2015).

[30] O. Russakovsky, *et al.*, "ImageNet large scale visual recognition challenge", *Int. J. Comput. Vision* 115(3), 211–252 (2015), doi: 10.1007/s11263-015-0816-y.

[31] VGG network weights, [Online], Available: https://www.robots.ox.ac.uk/~vgg/research/very_deep/

[32] Resnet network weights, [Online], Available: https://github.com/KaimingHe/deep-residual-networks.

[33] P. Leydon, M. O'Connell, D. Greene, K. M. Curran, "Bone Segmentation in Contrast Enhanced Whole-Body Computed Tomography", *arXiv* (2020), https://arxiv.org/abs/2008.05223.