On Efficient Extraction of Pelvis Region from CT Data

Tatyana Ivanovska[1], Andrian O. Paulus[1], Robert Martin[2], Babak Panahi[3], Arndt Schilling[2]

[1]Department for Computational Neuroscience, Georg-August Universität Göttingen
[2]Clinic for Trauma Surgery, Orthopaedics and Plastic Surgery, University Medicine Göttingen
[3]Institute for Diagnostic and Interventional Radiology, University Medicine Göttingen
tiva@phys.uni-goettingen.de

Abstract. The first step in automated analysis of medical volumetric data is to detect slices, where specific body parts are located. In our project, we aimed to extract the pelvis regionfrom whole-body CT scans. Two deep learning approaches, namely, an unsupervised slice score regressor, and a supervised slice classification method, were evaluated on a relatively small-sized dataset. The result comparison showed that both methods could detect the region of interest with accuracy above 93%. Although the straightforward classification method delivered more accurate results (accuracy of 99%), sometimes it tended to output discontinuous regions, which can be solved by combination of both approaches.

1 Introduction

For efficient analysis of medical images, it is beneficial to pre-extract the region of interest for further processing steps. This allows for the reduction of the amount of data, so that subsequent computations are boosted and possible errors are reduced. This is especially important if the number of slices comprising the region of interest is relatively small, when compared to the complete sequence size.

In the literature, the approach of labeling the anatomical regions is referred as bodypart recognition. It serves as an initialization module for anatomy detection or segmentation algorithms [1]. Classical approaches utilized extensions to Generalized Hough Transform (GHT) [2] for such a task. For instance, Seim et al. [3] followed the approach of Khoshelham [2] to localize the pelvic bone for further segmentation. However, Seim et al. stated that the performance of the GHT strongly depends on the quality of the template shape and the speed is rather low, since it is a brute force method. Since the ultimate goal of our project is to analyze data with multiple fractures and mislocated bones, such a method might fail in localization of the pelvis region.

Deep learning methods that do not require any prior knowledge of the data can be roughly separated into two categories: supervised classifiers, which are

trained on labeled images, and unsupervised regressors, where the anatomical landmarks are learned from the slice ordering. Roth et al. [4] used labeled CT scans from 1675 patients to train a convolutional neural network (CNN) to differentiate between 5 classes (neck, lung, liver, pelvis, legs). Yan et al. [1] proposed a multi-stage deep learning framework to discover local discriminative patches and build local classifiers to differentiate between 12 anatomical classes using data from 675 patients. Yan, Lu, and Summers [5] proposed an unsupervised body part regressor that constructs a coordinate system for the human body and outputs a continuous score for each axial slice, representing the normalized position of the body part in the slice. These so-called slice scores are learned from intrinsic structural slice ordering information in CT volumes. The main advantage of this approach is that it does not require labeled data for training.

For our application, the pelvis region detection is required in the first step, and other anatomical classes can be disregarded at the moment. We utilized whole-body computed tomography (CT) data [6] of $\approx$ 100 patients. As this was a relatively small dataset, the goal of our study was to explore the priority approach for extraction of pelvis with limited data. We analyzed two deep learning approaches. The first method is an unsupervised body-part regressor proposed by Yan, Lu and Summers [5]. The second method uses a classical VGG-based convolutional neural network [7] for slice classification task, which can be considered a simplified version of the approach by Roth et al. [4]. We applied both methods to our dataset, present the results, and analyze the findings.

2 Materials and Methods

2.1 Materials

The study was approved by the local ethics committee of University Medicine Göttingen (RefNo: 1-5-19). Anonymized whole-body CT data of 93 patients were included into the project.

Several example slices from different patients are presented in Fig. 1. The data were converted to a standardized slice thickness of 5 mm. The spatial resolution was 512×512 and the number of slices varied for each patient. The

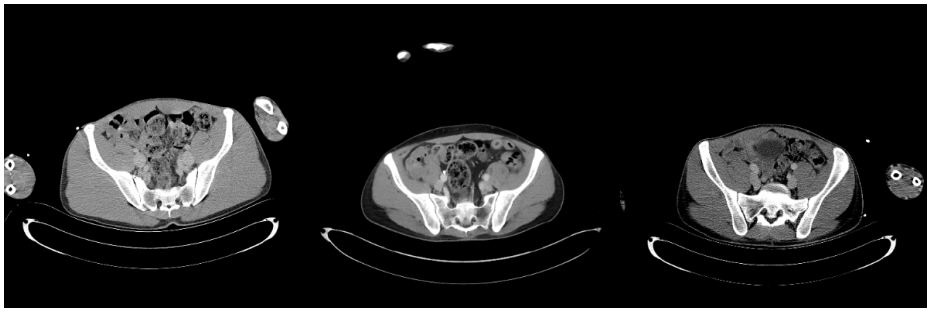

Fig. 1. Three example slices depicting pelvis region.

average slice number was 236 ± 56, and the median slice value was 234. It can be observed that pelvis region occupies 43 ± 2.2 slices, i. e. $\approx \frac{5}{6}$ of the slices can be potentially disregarded in subsequent processing steps.

The data were split into training, validation, and test sets patient-wise. 10 and 9 patients were put to the test and validation sets, respectively, the 74 scans were used for training. Pelvis region in each dataset was identified by an experienced observer.

2.2 Methods

Unsupervised Body Part Regressor is a novel learning method that learns the body part knowledge from inter-slice relationships in a self-organization process [5]. It has six convolutional layers from a standard network (VGG16), a global average pooling layer and a fully connected layer, which outputs the score. A novel loss function that consists of two terms serves for efficient learning. Yan, Lu and Summers collected around 800 CT data from 420 subjects. They also provided an open source implementation of their method [8]. We re-implemented the method in pytorch using the original source code and the implementation by G. Chartrand [9], which was trained and evaluated on the IRCAD dataset [10]. It was suggested in the original publication to reduce the width and height of slices. We have also observed in our experiments that usage of smaller input image sizes produces more accurate results. Therefore, the original data were resized to 64×64. The image intensities in Hounsfield units [11] were clipped to a range of $[-300, 300]$ to provide more contrast to soft tissues. Thereafter, intensities were scaled to a range of $[0, 1]$ and tripled in order to create a multichannel image. One third of the training volumes were augmented by shifting the individual slices to the left and right, as well as up- and downwards. The amount of shifting are randomly chosen for each individual volume and ranged from $[10, 70]$ pixels. Training runs were initialized using the same parameters found in [9], with the exception of the batch composition. Here, batches consisted of 12 different volumes each represented by 14 equidistantly spaced slices. After obtaining the scores, a histogram-based method was applied to compute the slice values for each region in a volume. For more details, we refer to the original work [5].

VGG-based Slice Classifier was built using VGG11 [7], the lightest model from the VGG family. The model consisted from 11 weight layers and 5 pooling layers: eight convolutional layers, each followed by a ReLU activation function, five max pooling operations, each reducing feature map by 2, and three fully connected layers. All convolutional layers had 3×3 kernels. The first convolutional layer produced 64 channels and then, as the network deepened, the number of channels doubled after each max pooling operation until it reached 512. On the following layers, the number of channels did not change.

We replaced the last classification layer to a layer for 2 class classification, and used the Imagenet weights. The standard augmentation tools, such as scaling, rotation, horizontal and vertical flips were applied to the data. The input image

Table 1. $\mu \pm \sigma$ metrics for validation and test sets for both approaches.

TP	FP	TN	FN	A	P	R	F1
Unsupervised Body Part Regressor							
Validation Set							
36 ± 6.4	7.2 ± 5.8	174 ± 57	7.6 ± 6	0.93 ± 0.05	0.84 ± 0.1	0.83 ± 0.1	0.83 ± 0.1
Test Set							
34 ± 4.9	4.8 ± 7.1	174 ± 44.5	9.8 ± 4.8	0.93 ± 0.05	0.9 ± 0.1	0.78 ± 0.1	0.82 ± 0.08
VGG Classifier							
Validation Set							
42.6 ± 3.1	0.7 ± 0.9	187 ± 57.4	0.2 ± 0.4	0.99 ± 0.02	0.98 ± 0.02	0.99 ± 0.01	0.99 ± 0.01
Test Set							
43.4 ± 1.8	0.3 ± 0.5	179 ± 41	0.3 ± 0.5	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.01	0.99 ± 0.08

size was $224 \times 224 \times 3$. CT data preprocessing steps were the following: the intensities in Hounsfield units were clipped to a range of $[0, 3000]$; each volume was z-score normalized. To make the inputs compatible with the VGG architecture, the original images dimensions (512×512) were resized to 224×224, and intensities were scaled to a range of $[0, 1]$, the images were copied 3 times to be used as a multichannel image, and then normalized using $mean = [0.485, 0.456, 0.406]$ and $std = [0.229, 0.224, 0.225]$.

2.3 Evaluation Measures

The following values were used for evaluation: true positives (TP), false positives (FP), true negatives (TN), false negatives (FN), which refer to correctly and incorrectly classified slices, where pelvis region is located. From these values the following metrics were computed: Accuracy $A = \frac{TP+TN}{TP+TN+FP+FN}$, Precision $P = \frac{TP}{TP+FP}$, Recall $R = \frac{TP}{TP+FN}$, and F1 score $F1 = \frac{2TP}{2TP+FP+FN}$.

3 Results and Discussion

The unsupervised regressor was trained for 50 epochs with the following parameters: batch dimensions $g = 12$,$m = 14$, Adam optimizer with $lr = 0.0001$. The VGG classifier was trained starting from ImageNet weights for 20 epochs using Adam optimizer with $lr = 0.0001$ and batches of 64 images. The results are summarized in Table 1.

As it can be observed in Table 1, the VGG classifier produced highly accurate results with maximally two false positive and one negative slices per patient. In most cases misclassifications appear at the beginning or the end of pelvis region. An illustrative example is presented in Fig. 2. Actually, the errors of the predictions (left and right slices in Fig. 2) contain some minor parts of the pelvic bones, and thus, the expert had possibly not always marked them for all datasets consistently. We leave the investigation of inter- and intra-observer

variability for future work. The standard deviation in TP and TN is explained by the fact that the number of slices occupied by pelvis region differed for each patient.

Although the unsupervised regressor was able to find the correct regions as presented in Fig. 3, the results had significantly lower rates than by the supervised classification method (cf. Table 1). This is probably explained by the fact that we used about 10 times less data for training, when compared to the original publication, namely, 73 vs. 800 volumes.

However, the classification-based approach suffered from one drawback: since the method considers only 2D slices and no 3D information is taken into account, it can potentially produce disjoint regions, e. g., miss some slices within or detect something outside the region of interest. For instance, the classifier in some rare cases detected several slices in the shoulder region and marked them as pelvis, since it visually reminded the lateral parts of pelvic bones. The probability

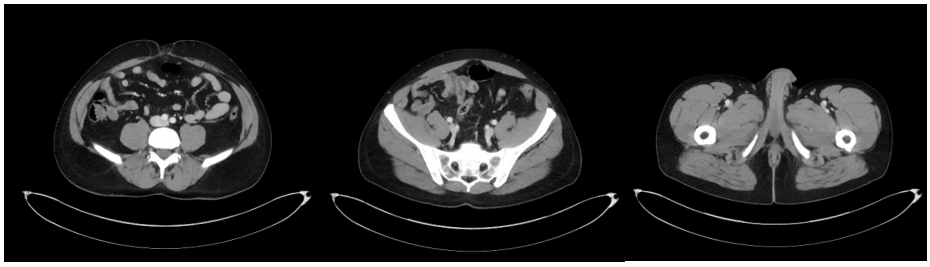

Fig. 2. Predictions of the VGG-based classifier. Left: false positive slice at the beginning of pelvis; middle: correct prediction; right: false negative slice at the end of pelvis.

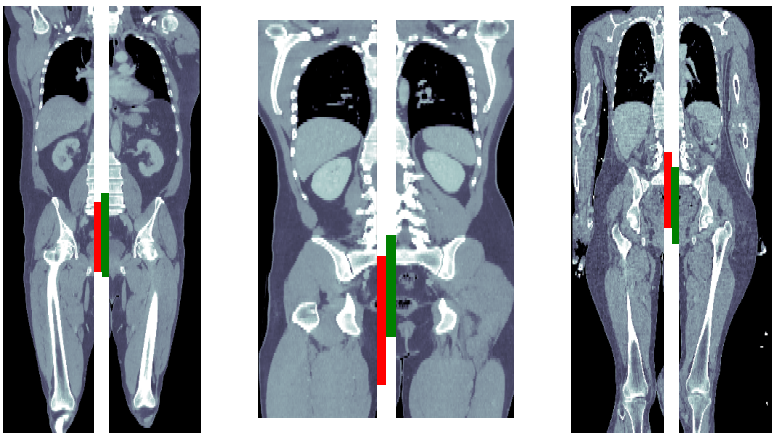

Fig. 3. Results of the unsupervised regressor detection overlaid with data from three patients are shown. The algorithm detection of pelvis region is marked in red, the expert readings are shown in green.

values of the classification were similar to the ones at the end of the pelvis region and no straightforward thresholding was possible. This problem could be solved by using a combined approach, where the results from both methods are considered together, since the unsupervised method produces one continuous region by definition. We plan to add more training data to the unsupervised regressor and leave the extension for future work. Moreover, we also plan to investigate, how well such an approach would perform, when more regions in the body must be localized.

4 Conclusion

In this study we compared two deep learning approaches for the task of body part detection: an unsupervised regressor and a slice classification network, to label pelvic bones in whole-body CT data with a relatively small dataset. Even with this small amount of data, the accuracy was already high. This suggests, that given more training data and combining both methods, such an approach will presumably lead to efficient and reliable reduction of data for analysis of CTs of the pelvis.

References

1. Yan Z, Zhan Y, Peng Z, et al. Multi-instance deep learning: Discover discriminative local anatomies for bodypart recognition. IEEE transactions on medical imaging. 2016;35(5):1332–1343.
2. Khoshelham K. Extending generalized hough transform to detect 3d objects in laser range data. In: ISPRS Workshop on Laser Scanning and SilviLaser 2007, 12-14 September 2007, Espoo, Finland. International Society for Photogrammetry and Remote Sensing; 2007. .
3. Seim H, Kainmueller D, Heller M, et al. Automatic Segmentation of the Pelvic Bones from CT Data Based on a Statistical Shape Model. VCBM. 2008;8:93–100.
4. Roth HR, Lee CT, Shin H, et al. Anatomy-specific classification of medical images using deep convolutional nets. In: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI); 2015. p. 101–104.
5. Yan K, Lu L, Summers RM. Unsupervised body part regression via spatially self-ordering convolutional neural networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). IEEE; 2018. p. 1022–1025.
6. Furlow B. Whole-Body Computed Tomography Trauma Imaging. Radiologic Technology. 2017;89(2):159CT–180CT.
7. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556. 2014;.
8. Original implementation by Yan, Lu, Summers;. https://github.com/rsummers11/.
9. Implementation in Keras by Gabriel Chartrand;. https://github.com/Gabsha/ssbr.
10. IRCAD Dataset;. https://www.ircad.fr/research/3dircadb/.
11. DenOtter TD, Schubert J. Hounsfield Unit. In: StatPearls [Internet]. StatPearls Publishing; 2019. .