Automated Shoe Metrology by X-ray Computed Tomography

Martin LEIPERT *[1,2], Gabriel HERL [1], Michael MÜLLER [3], Joshua MESKEMPER [3], Simon ZABLER [1]
[1] Deggendorf Institute of Technology, Deggendorf, Germany;
[2] Pattern Recognition Lab, FAU Erlangen, Erlangen, Germany;
[3] OneFid, Cologne, Germany

https://doi.org/10.15221/23.48

Abstract

Automation of shoe metrology is crucial to providing fit information for shoes on a large scale. Here, we examine a segmentation technique to extract the inner shoe volume (ISV) from Computed Tomography (CT) data—the proposed approach leverages artificial neural networks to extract shoe parts for automated metrology precisely. The neural network architecture is customized to facilitate the extraction of ISV by integrating spatial attention mechanisms. Furthermore, a neural network segmentation algorithm removes filler materials virtually. This process yields enhancements of 1.3% in F_1-score through material removal and an additional 1.4% through the incorporation of spatial attention. Notably, spatial attention mechanisms yield improved outcomes at the aperture of the shoe. The elimination of filler materials reduces false positive segmentations. The segmentation outcomes are utilized to generate surface meshes. These are compared to surface meshes derived from annotated data. We measure an average Hausdorff distance between annotated and labelled data of 2.1 mm. The discrepancy is primarily attributed to deformations and artifacts. On both sets, we measure the effective shoe length. Precision and accuracy metrics for the extracted measurement from ANN-segmented data attain 0.8 mm and 1.8 mm, respectively. For meshes obtained from label data, the precision is 0.2 mm, and the accuracy is 2.5 mm. Our findings underscore the accuracy of the extracted shoe interior volumes, rendering them suitable for metrological applications. Limitations include unsolved issues with separation reliability and deformation.

Keywords: Automated Shoe Fitting, Shoe Metrology, Computed Tomography

1. Introduction

The importance of well-fitted shoes cannot be overstated, as they significantly influence foot health and overall well-being. Unfortunately, the current practice of choosing and recommending shoes relies solely on standardized shoe sizes. The latter and width measurements often lead to a prevalence of ill-fitting shoes among the population. This issue is especially concerning for individuals with medical conditions such as diabetes, where optimal footwear is crucial to prevent complications. Buldt et al. conducted a comprehensive review of studies investigating foot-shoe fit. They found alarming rates, ranging from 63% to 72% of participants wearing shoes that do not match their measured optimal shoe size and width [1]. Buldt et al. concluded that there is a need for a standardized system of shoe sizing that considers natural variations in shoe and foot morphology. Stanković et al. explored the three-dimensional shape variations of healthy individual feet, identifying six crucial measures encompassing 92.59% of the total foot shape variation and are located mainly at the forefoot [2]. These essential measures, including arch height, combined ball width and inter-toe distance, global foot width, hallux bone orientation (valgus-varus), foot type (e.g., Egyptian, Greek), and midfoot width, are inadequately covered by the conventional sizing system.
Additionally, the development of e-commerce further drives the demand for an improved sizing system, as there is no possibility to try shoes before ordering [3]. This results in high return rates and customers wearing ill-fitting shoes. The German Institute for Standardization introduced an improved sizing system with the DIN SPEC 91416 norm that defines an array of metrics for the shoe, shoe lasts and the corresponding anatomy of feet. Three-dimensional foot scanning with laser scanners is well-researched and can obtain metrics like foot length, ball of foot length, outside ball of foot length, foot breadth diagonal, foot breadth horizontal and heel breadth with high precision and robustness [4]. However, the metrology of shoes is still in its infancy. Optical metrology is impractical since it requires unpacking the shoes and because the inner dimensions, which are most important, are difficult to access. Automated 3D metrology of packed shoes is feasible using Computed Tomography (CT); the digital segmentation of this data is the core of this report.

* e-mail: martin.leipert@fau.de

1.1. Related Work

For shoes, there is often no corresponding fitting data available. There are four reasons for this: first, shoe design is mainly focused on fashion and not on fit. Second, the morphological fit descriptors of shoes are usually unknown. Third, shoe lasts used for manufacturing are rarely available. Fourth, shoe lasts change during the manufacturing process due to erosion. Hence, precise three-dimensional metrology of shoes is needed. There is no gold standard for dimensional metrology of shoes, and various methods are under investigation. Revkov and Kanin investigated the efficacy of tactile probes for measuring the inner volume of shoes and generating three-dimensional meshes. Their findings revealed an acceptable level of accuracy, with deviations of approximately 1 mm from the ground truth. Additionally, the method is also able to capture deformation and stretching behavior [5]. Another approach by Omrcen and Jurca involves filling shoes with a contrast agent and capturing two-dimensional X-ray projection images to obtain a 3D shoe last [6].

Computed Tomography (CT) has emerged as a valuable tool. Kuper et al. utilized a medical CT scanner to measure unboxed shoes, extracting their inner volumes and obtaining various metrics [7]. Jo and Park applied CT to measure the intrinsic dimensions of firefighter boots. They used the extracted inner volume directly to fit it to the measured feet surfaces of firefighters; they demonstrated that feet and boots mismatched significantly [8]. The possibility to scan shoes on a large scale, automatically and without unboxing them, makes CT interesting for a commercial measurement process. In e-commerce, shoes are obtained from suppliers, and their dimensions are largely unknown except for the shoe size and sometimes shoe width. For these, a variety of standards exist, like the Paris Point shoe size metric in continental Europe. Customers often order multiple pairs of shoes of different sizes and return those that do not fit. Better fitting algorithms and exact recommendations could reduce return rates significantly. Wittmann et al. developed an algorithm for this use case to digitize the shoe's inner volume and extract its surface mesh from boxed half shoes based on region growing and active contours [9].

Recently, the authors improved the segmentation of the inner shoe volume (ISV) from CT scans of boxed shoes using an Artificial Neural Network (ANN) [10]. Trained on a small set of different shoes, the latter displays high generalization abilities and yields an F_1-score of 81 % for ISV extraction. The main problem is filler material inside the shoe, mostly packing paper, that is not segmented correctly as ISV. Filler material provides a disturbance to the ANN: as convolutional kernels are used; the texture of the filler material results in a feature response. The network needs to learn that this texture is part of the ISV; hence, this filler material increases the complexity of the segmentation problem. Another mayor problem found is the space between two shoes; if these are close together, the space between shoes is often mistaken for ISV. Minor problems include that the corners of the packing are sometimes mistaken for ISV and that the border between shoe ISV and background is not clearly defined at the shoe opening. As ISV and background are both mostly air-filled spaces, that is only differentiable by the context.

1.2. Contribution

The present work features automated ISV extraction from a batch of boxed half shoes based on the method developed in [10]. As in the original work, the F1-score for the extraction was only 81 %; the ANN is improved by introducing spatial attention to the extracting part of the network and by automated removal of filler material. With further post-processing, the voxel (three-dimensional pixel) wise prediction of the ISVs is turned into surface meshes. These are evaluated against meshes extracted from ground truth data (manual segmentation) and compared to the standardized sizes of the shoes. Additionally, we generally examine the influence of filler material, deformation, and artifacts on the ISV. We propose strategies for compensating for these factors and hence avoid possible errors during the extraction of shoe metrics from the DIN SPEC 91416 norm.

2. Materials and Methods

The future goal of this work is to extract metrics from the DIN SPEC 91416 norm via CT. This norm defines a comprehensive set of 33 measures for feet, shoe lasts, and shoes. Most of these metrics can be extracted from the inner shoe volume, including standard measurements such as shoe length (shoe size) and anatomic ball girth (shoe width). Further girth and width measures cover the entire foot from heel to hallux. Hence, a precise extraction of the ISV is required, especially at the forefoot.

2.1. Computed Tomography

In our research, we utilize industrial Computed Tomography (CT), a method for measuring the three-dimensional X-ray opacity of objects. This property corresponds to the object's density, as visible in Figure 1, where a CT slice image of a sneaker is shown. Dense parts absorbing more radiation appear white (e.g., outsole), less absorbing parts (insole) grey, whereas air, absorbing no radiation, appears black. All images here have low grey-value enhancement; otherwise, only heavily absorbing parts (metal, dense sole) would be well visible. For readers unfamiliar with CT, we recommend the following literature: the introduction [11] and principles chapter of the textbook Industrial Computed Tomography [12]. Computed Tomography is prone to artifacts. These are discrepancies between the measured and actual attenuation of objects. As shoes are multi-material objects, metal and beam hardening artifacts are of relevance. Therefore, we recommend the corresponding textbook chapter about artifacts [13]. The problem of artifacts is demonstrated in Figure 2, where an apparent streaking around a metallic splint is visible.

Figure 1 CT-scan of a boxed and squeezed sneaker.

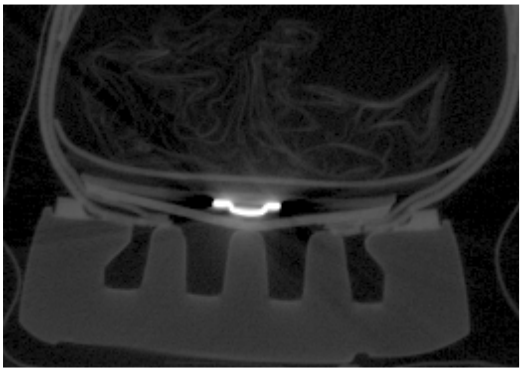

Figure 2 Streaking and shading artifacts at stabilizing metal splint (white) in a shoe. Above the splint filler material, in this case packing paper, is visible.

2.2. Extraction of the Inner Shoe Volume

In this study, the ISV is extracted from shoe CT scans by two distinct methods, which are common when machine learning is employed for segmentation:

A. **Manual Annotation:** The first approach involves manually annotating the raw CT data to segment the ISV in Slicer 3D [15]. An expert user carefully identifies and outlines the boundaries of the ISV and other shoe parts in the CT scans for annotation. They are then double-checked. The generated data are regarded as ground truth, the best achievable segmentation out of the CT volume.

B. **Automated Segmentation with ANN:** The second approach utilizes an Artificial Neural Network (ANN) for automated segmentation. This approach is explained further in the following section.

The ANN-based approach is an improved version of the approach presented in [10]. There, the ANN takes the grey value volume of the CT scan as input and predicts a probability map of each voxel belonging to a single class. For the segmentation of ISV, the algorithm only distinguishes the classes ISV and background. The main problem of this segmentation is that the ISV is mostly air-filled space that is sometimes enclosed by the shoe, while most of the background is air-filled space around the shoes. Hence, there is a border definition problem at the shoe opening, with no clearly defined boundary between in- and outside.

The shoe is often stuffed with filler material; the area with filler material also belongs to the ISV. Thus, the filler material should be segmented as ISV. Two types of filler material dominate in our experiments: packing paper, which is of low density and has a characteristic texture. The second type is carton inlays in shoes. In contrast to packing paper, this material is harder to distinguish from shoe material, as the carton is of similar thickness and density as the shoe upper, as shown in Figure 4. It also often is in direct contact with the shoe upper. In the previous experiments, we observed problems with the ISV prediction to segment an area that contains filler material [10]. Additionally, there were false positive segmentations of ISVs in the corners of packages. However, the results of the first segmentation stage indicated that our ANN is also well able to segment packing material and the carton [10]. This is also shown in Figure 3, where the packing paper and the carton are correctly segmented. Therefore, we propose to use this segmentation result to virtually remove carton and filler material. The removal of the filler material and carton is done by taking the segmentation result from

the previous step, respectively, the blue area from Figure 3. There, the filler material and carton area in the input volume are replaced by zero values. These volumes without the carton and filler material are now the input data for the segmentation algorithm.

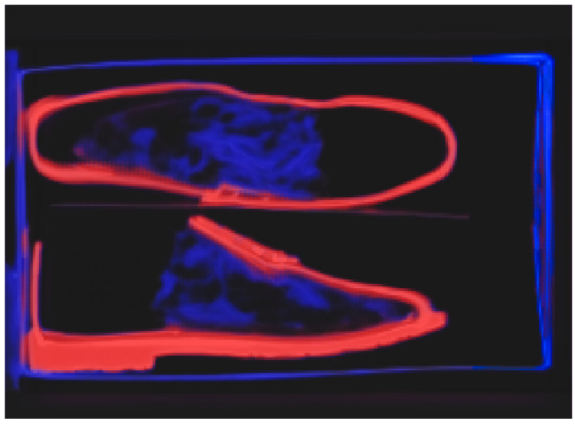

Figure 3 Segmentation result for a shoe filled with paper by our ANN from [10]. Voxels labelled shoe in red, filler material and carton labelled blue.

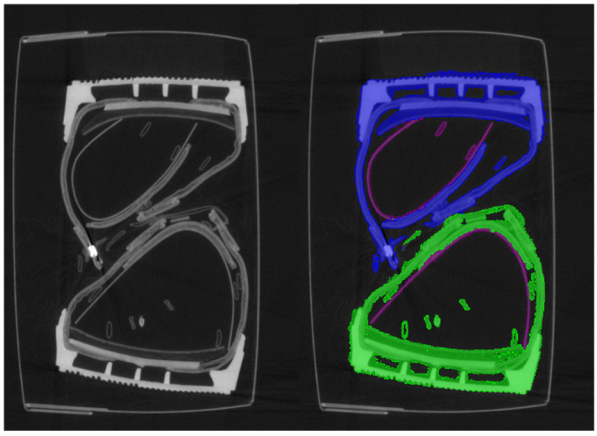

Figure 4 Left: Raw grey value image. The filler carton is hard to distinguish from the shoe material. Right: Corresponding labeling with shoe in blue and green. Filler material in magenta.

Another problem from the previous segmentation is the area enclosed between two shoes. This air-filled space is nearly enclosed by shoes. So, this area, for the network, resembles more the ISV; however, it is background. Therefore, we propose the usage of spatial attention, so a feature weighting mechanism for different regions in the image [13]. This is a special feature weighting and is based on the alignment of the weights used in the UNet. With the Residual-Squeeze-And-Excitation blocks, we already employ a feature weighting algorithm that performs channel-wise feature weighting. The feature weighting is computed in every block by global classification on the mean of all feature maps in the prediction. In contrast to this, spatial attention is computed in the decoding path of the UNet. It combines the feature map from the previous layer and the feature map from the corresponding encoding layer to a regional weighting. This regional weighting is then multiplied to the feature map from the encoding path. So spatial attention suppresses or highlights regions in the feature maps coming from the encoding path. In consequence, the expectation here is to suppress the area between two shoes and get a clearer border for the inside and the outside of the shoe ISV.

Figure 5 Half shoe model from our dataset.

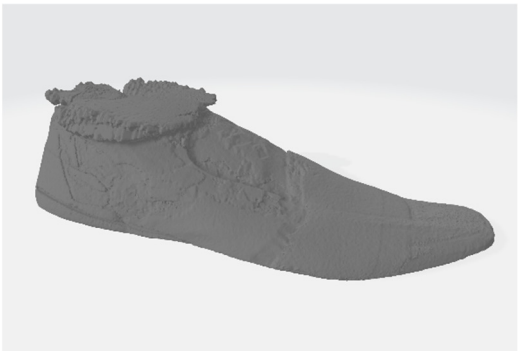

Figure 6 3D rendering of ISV generated automatically by the ANN (for the shoe from Figure 5).

Further postprocessing is required to turn a predicted probability map into a (iso-)surface mesh. The postprocessing as implemented here is depicted in the sketch in Figure 7. Watershed segmentation separates the single ISVs [15]. The distance transform serves to find seed points for the watershed transform. Binary operations are used for smoothing the binary masks and for filling holes. The marching cubes algorithm is used to obtain an isosurface from the probability map [15]. Finally, Laplacian smoothing is used to improve the meshes for metrological applications.

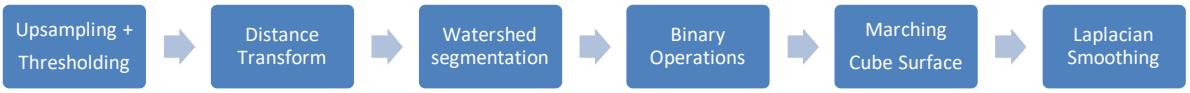

Figure 7 Image-processing Pipeline used for extracting and improving the ISV mesh from the ANN predictions.

2.3. Shoe Dataset

This study uses a dataset that compromises 24 half-shoes. Each pair of shoes is scanned in their shoebox with standard industrial X-Ray CT systems; the scanning parameters are depicted in Table 1. For CT reconstruction, we use the Feldkamp David Kress (FDK) algorithm [18]. For the deep learning pipeline, we split the set into fifteen training, three validation and six test pairs.

Table 1 X-Ray acquisition parameters

Parameter	**Value**
Voltage [kV]	250
Current [µA]	6000
Detector matrix [voxel]	1936 x 1936
Voxel size [µm]	170.9
Pixel size [µm]	200
Exposure time [ms]	330
Projections [number]	1600
Focus object distance [mm]	2000
Focus detector distance [mm]	2340
Prefiltering [mm]	1 Cu and 0.5 Sn

3. Results and Discussion

To evaluate our improvements in the ISV extraction, we train the network on the shoe dataset as described before. We use four settings:

(1) The baseline setting. This is the best-performing setting from the previous work in [10]. The used model is the Residual-SE UNet. It uses data that still contain carton and filler material to train and predict.
(2) This setting introduces spatial attention to the Residual-SE UNet. The data for training and prediction contain filler material.
(3) This setting trains the Residual-SE UNet without attention to data, where carton and filler material are removed virtually.
(4) This setting now uses the Residual-SE UNet, including spatial attention, and trains on data with removed carton and filler material. So, it combines the improvements from setting (2) and (3).

The training parameters are mostly identical to the training in [10]: For the Residual SE UNet, with and without attention, 16 feature maps are used in the first layer of the network model. The network is trained with weighted Cross-Entropy-Loss for 120 epochs and an initial learning rate of 0.002. In contrast to the original work, we employ AdamW for optimization instead of SGD. We use data augmentation techniques as described in [10] and additionally apply translation. Data augmentation happens on the fly and with random parameters, so the samples differ between each training epoch.

To evaluate the segmentation result, the F_1-score is used. This is the harmonic mean of precision and recall. This metric is suitable for class imbalanced data as they occur here, where most voxels are of class background. To assess the accuracy of the ANN-based mesh extraction, we calculate the Hausdorff distance between the meshes from our ANN prediction and the meshes generated from label data. The comparison is conducted in the forefoot area of the meshes, the most relevant for fitting according to DIN SPEC 91416 and the problematic shoe opening is excluded. Additionally, the meshes are compared visibly. To assess the quality of the generated meshes for fitting, we extract the only available metric from the manufacturer, the effective shoe length. We compare the metric for the meshes from the label and from predicted data to assess if the quality of the predicted meshes is comparable to label data for fit extraction.

3.1. Improvements of the ISV extraction

Tab. 2 lists the F_1-scores (harmonic mean of precision and recall) for the half-shoe dataset and the different improvements introduced here. Without removing filler material, introducing spatial attention improves the F_1-scores by 1.4 % for ISV. The filler material removal improves the F_1-score for ISV by 1.3 % compared to the baseline. In the combined version, the score improves by 4.3 % compared to the baseline. In summary, both approaches lead to independent improvements that also work in combination. While removing filler material simplifies the task, the attention includes spatial context in the prediction.

Table 2 F1-scores for the different setting of the segmenting ANN used here.

Setting	Network	Filler Material	$F_{1,\text{ Inner Shoe Volume}}$	$F_{1,\text{ Background}}$
1	Residual SE	Not Removed	80.2 %	98.5 %
2	Residual SE + Attention	Not Removed	81.6 %	98.6 %
3	Residual SE	Removed	81.5 %	98.6 %
4	Residual SE + Attention	Removed	84.5 %	98.8 %

Figure 8 contains the predicted probability maps for ISV, three shoes, and four settings. The settings, including attention, improve the prediction at the shoe opening, except for squeezed shoes, as in the case of the third volume. Spatial attention improves the prediction at the shoe opening, demonstrated for the 2nd and 4th setting for the 1st and 2nd shoe volume in Figure 8. Also, for shoes with filler material, like at the tip of the first shoe volume, the introduction of spatial attention improves the segmentation outcome. The same hold for filler material removal, which is well visible at the first volume for the first and third setting. As the introduction of spatial attention adds additional layers and, therefore, additional parameters, we assume that with a more extensive training set, the usefulness of spatial attention improves even further.

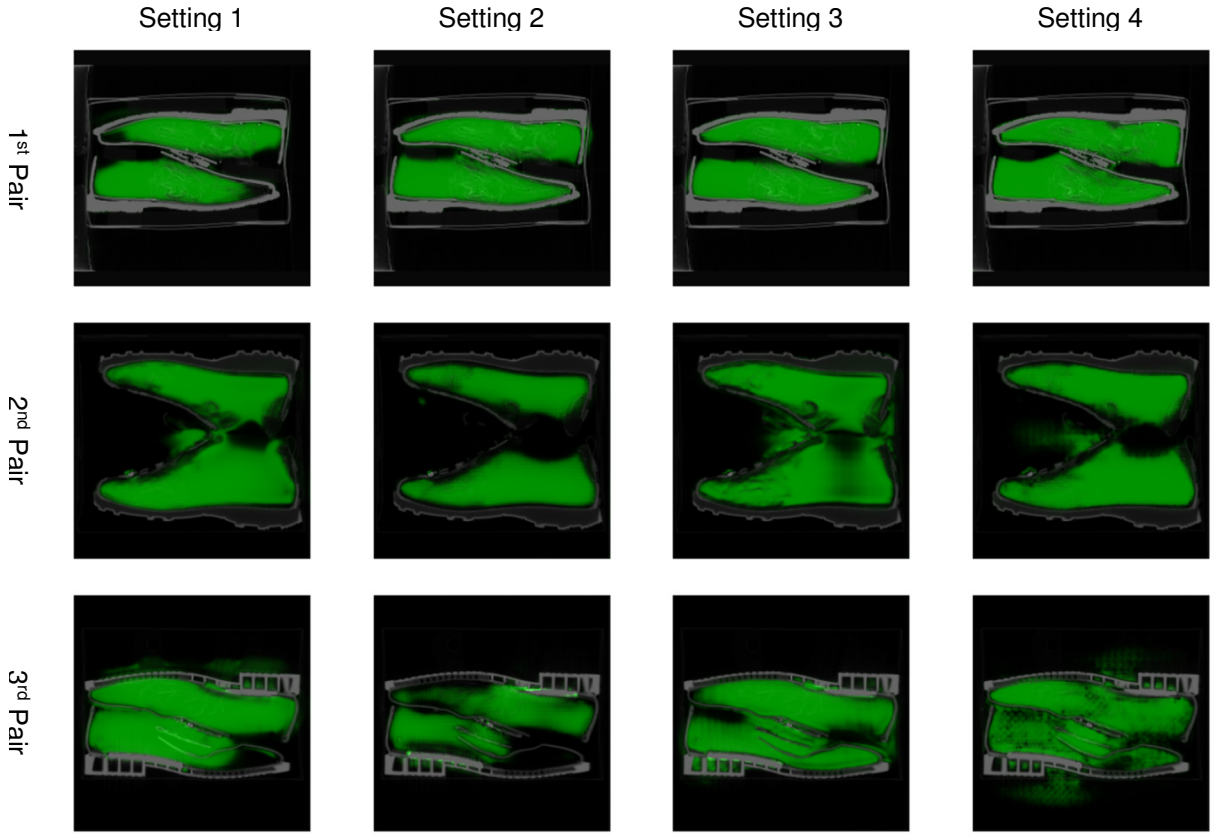

Figure 8 ISV predictions for three different shoes and the single settings. The settings including attention improve the prediction at the shoe opening, except for squeezed shoes.

Suppose two shoe openings are in close vicinity. In that case, the ISVs are directly connected, as for the second shoe in Figure 8, and this poses a problem to the separation with watershed segmentation. Also, the segmentation artifacts between two shoes are often connected to the opening, leading to

erroneous segmentations. The latter is especially visible in the third setting for the second volume, where the ISVs are nearly connected at the shoe opening. To further increase the accuracy at the opening and evade problems with the area between shoes being predicted as ISV, we propose the usage of panoptic segmentation instead of the watershed, such that single shoes may be segmented [15]. Additionally, a segmented single shoe can be oriented uniformly, which improves the segmentation results, as in the case of the insole, outsole, and shoe upper segmentation [10]. As visible in Figure 8 for the first two pairs of shoes, the filler material is also beneficial for the result as there is less deformation. The third pair of shoes is deformed, and the space between both shoes is narrow; the segmentation result allows not to draw a clear boundary as for the 1st and 2nd pair. Also, there are false positive areas outside the shoe for the 4th segmentation setting.

3.2. Accuracy of the Virtual Shoe Lasts

We first evaluate the ISVs' accuracy by comparing the segmented ISV meshes (after post-processing) to the ISV meshes generated from label data for all pairs of shoes from the test set. These ISVs serve as virtual shoe lasts, from which we want to take fitting measures in the future. To evaluate the accuracy of the shoe lasts, we take predicted meshes from the fourth setting and compare these to ground truth shoe lasts. This is done by calculating the Hausdorff distance between both meshes at the forefoot until the shoe opening. This area is of the most relevance for fitting. Additionally, we evade the problem that the mesh border at the opening is unclear. An additional comparison is done for the ISV segmentations by visualizing the difference between ground truth and predicted ISV as in Figure 11. There, the deviation of the segmentation border is mostly within 1 to 2 voxels, 0.33 mm to 0.66 mm between both surfaces. There is a deviation of smaller than one voxel for significant parts, these are marked in grey. The exception is for the shoe opening, where the approach segments significantly more area than it was labelled in the ground truth.

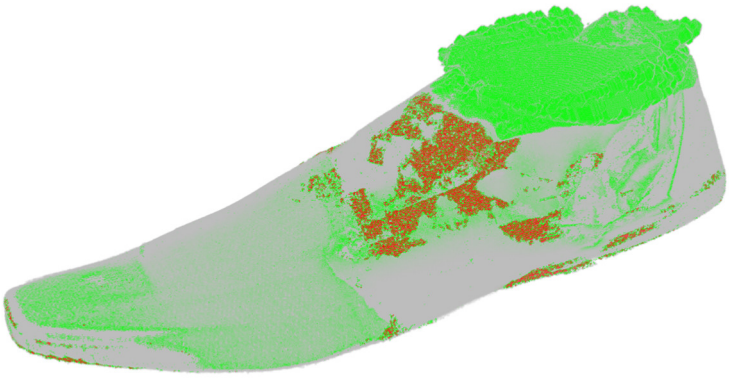

Figure 9 Voxelized difference of predicted volume and ground truth data for one ISV segmented with Setting 1. Grey indicates matching surface areas. Green indicates a prediction that is not in the ground truth data, red labels vice-versa. As visible, the main difference is at the opening of the shoe, while for the rest the deviations are minor and ranging below one millimeter.

The Hausdorff distances are computed from the faces of the output meshes. In Table 3, the results for the test set are shown. The prediction and reference range mean deviation between 1.1 mm and 3.6 mm. At the same time, the standard deviation is between 1.7 mm and 5.4 mm. In comparison, the difference, e.g., between two shoe sizes (in Paris Point), is 6.67 mm.

The standard deviation score ranges between 1.7 and 5.5 mm. This and the maximum values, which are ten times the value of the mean, indicate that there are measurement areas where there is a high deviation between the ground truth mesh and the predicted ISV mesh. An example is shown in Figure 10, where there are two main areas of outliers that explain these large deviations: false positive predictions outside the shoe and thin ISV areas. False positive predictions are sometimes connected to the correct ISV via a small opening, e.g., between the tongue and shoe upper; thus, they are included in the ISV prediction. The second case is false negatives in thin areas. These occur due to down-sampled volumes, where thin areas with a diameter of a few voxel diameters are hard to predict. These thin areas, e.g., are located between a loose insole and the fixed midsole or the gap between the insole and shoe upper.

Table 3 Hausdorff distances between label data and predicted meshes.

Pair	Shoe	Mean [mm]	Std. Dev. [mm]	Max [mm]
1	1	1.5	2.8	22.8
	2	1.1	1.7	13.0
2	1	2.1	3.3	15.9
	2	1.9	3.0	21.6
3	1	1.9	2.9	12.8
	2	2.9	4.5	29.2
4	1	1.3	1.9	11.1
	2	1.7	2.6	12.1
5	1	2.6	3.6	17.4
	2	2.3	3.4	17.3
6	1	2.2	3.4	22.8
	2	3.6	5.5	24.3

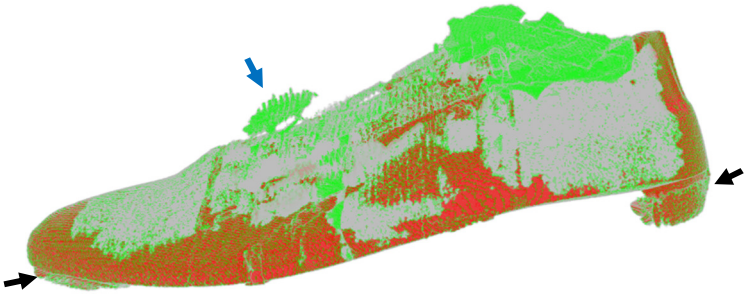

Figure 10 Shoe with cavity between insole and mid-sole. This also leads to problems for the prediction of the ISV. The gap, marked with black arrows, is not predicted correctly by the ANN. Also, there is an outlier, marked with a blue arrow, that is connected to the ISV by a small connection. Large areas however are marked grey and, in these areas, there is no deviation.

In Figure 12, we show the difference between the prediction and label data of deformed sneakers; the sneaker consists of soft material squeezed by the weight of the other shoe lying above. Figure 11 is the corresponding CT slice image. Here, the toe height is reduced to nearly zero. Especially where the gap is small, the prediction deviates from the ground truth data. This is partially also due to down-sampling, as the network cannot predict small gaps. Deformation, therefore, is an issue, especially as most fitting measures are taken at the forefoot, which is problematic if the forefoot is deformed. Deformation is mitigated by filler material, which we can remove virtually, as demonstrated before. Our suggested solution is the application of an energy functional on the inner volume that enforces a smooth outline of the shoes inner volume and "relaxes" the deformation.

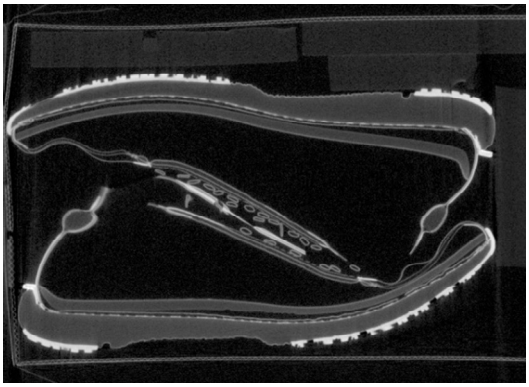

Figure 11 CT slice image of deformed sneakers.

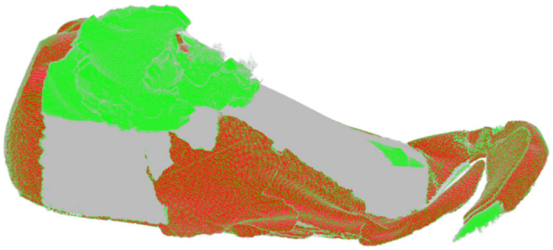

Figure 12 Segmented ISV of the volume from Figure 11. The segmentation misses some deformed areas.

Another source of deviations are artifacts. The shoe in Figure 13 contains a metallic zipper that causes streaking due to beam hardening artifacts. The zipper pattern is also visible on the extracted yet unsmoothed mesh in Figure 14. Metallic objects are frequently a part of shoes; they occur as zippers, splints, and eyelets. Options to correct these include the end-to-end removal of artifacts with neural networks [21]. Our neural network implicitly corrects metallic artifacts during segmentation if the inner volume is labelled correctly. However, due to the usage of thresholding in post-processing, the artifacts are reintroduced into the final mesh. This technique is usually sufficient in simple cases that preserve the original signal [22]. In more severe cases, iterative reconstruction techniques [22] and physics-based reconstruction can be used to correct artifacts [19].

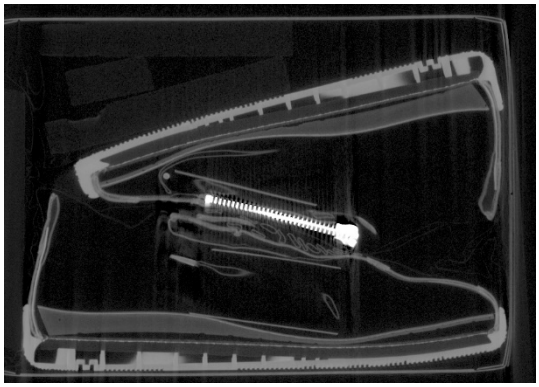

Figure 13 Shoe with a zipper and resulting streaking above and below.

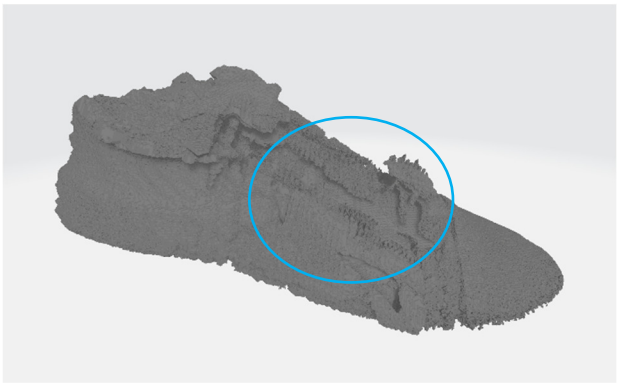

Figure 14 Resulting (still unsmoothed) mesh with clearly visible pattern from streaking cause by the zipper, marked by the blue circle.

3.3. Measurement extraction from Virtual Shoe Lasts

Finally, we extracted one measure from the ground truth data, that is the effective shoe length that to the shoe size. We took the six shoes from our test set and converted the shoe size into the metric effective shoe length. Shoe lengths were measured manually from the generated ISVs both for the label ISVs and the predicted ISVs. The mean, median and standard deviations are compared in Table 5. The measurements for single shoes are listed in Table 4. For the predicted data the mean deviation is 0.8 mm, the standard deviation is 1.8 mm. For the label data surfaces the standard deviation is 0.2 mm and the standard deviation is 2.5 mm. Both deviations are of comparable dimension, which indicates that the quality of the prediction for measurement extraction is comparable to the label data.

Table 4 Deviation of the effective shoe length for the label and predicted ISVs for pairs of shoes from our test subset. The data were obtained from half shoes of size 38 to 46.

Pair of Shoes	Effective Shoe Length [mm]	Deviation for the Shoe Length for Label Data ISV [mm]		Deviation of the Shoe Length for ANN Predicted ISV [mm]	
		1st shoe	2nd shoe	1st shoe	2nd shoe
1	266.7	2.3	0.3	-0.3	0.7
2	280.0	-0.1	-1.6	-0.3	-3.8
3	286.7	-5.7	0.2	-5.2	-1.2
4	293.3	-2.3	-0.7	-2.0	-0.6
5	293.3	0.9	-1.1	0.9	2.0
6	253.3	1.7	3.9	-0.3	0.0

Table 5 Deviations of the measures obtained from to the measures given by the manufacturers.

Label Data ISV			ANN Predicted ISV		
Mean Deviation Shoe Length [mm]	Median Deviation Shoe Length [mm]	Std Deviation Shoe Length [mm]	Mean Deviation Shoe Length [mm]	Median Deviation Shoe Length [mm]	Std Deviation Shoe Length [mm]
-0.2	-0.0	2.5	-0.8	-0.3	1.8

4. Conclusion and Outlook

We conclude that removing filler material and introducing spatial attention improve the result of ISV extraction with an ANN by 4.3 % compared to our Residual SE UNet baseline. Our approach can segment the ISV with sub-millimeter precision for most of the shoe surface, as we demonstrated in comparing segmented and label meshes in Figure 9. From these meshes, we can obtain the shoe size measure with a precision of 0.8 mm and an accuracy of 1.8 mm. However, the deviations above one millimeter in our results from Table 3, the deformed volume in Figure 12 and the zipper pattern at the mesh in Figure 14 indicate two influence factors that cause deviations: deformation and imaging artifacts. Metallic parts mainly cause imaging artifacts in shoes; solutions to correct these artifacts exist. We recommend applying direct reduction methods or iterative reconstruction, as artifacts present in shoes are considered moderate in the worst case [21]. Although we can segment deformed ISVs, these are unsuitable for shoe metrology; thus, further investigation is required to correct these deformed ISVs is required. We also found filler material beneficial in shoe segmentation, as our algorithm can remove it while it mitigates deformation. Another remaining problem is the border definition for the ISV; this is improved by spatial attention, as visible in Figure 8; however, the separation of shoes probably resolves this issue better. Once problems with the shoe opening, deformation and metal artifacts are solved, automated segmentation by ANN and metrology of the ISV for computational fitting by CT is sufficiently precise and accurate for industry and field application.

Acknowledgement

Our project was financed by the German Federal Ministry for Economic Affairs and Climate Action via Zentrales Innovationsprogramm Mittelstand (Central Innovation Programme for small and medium-sized enterprises) under grant number 16KN075740.

References

[1] A.K. Buldt, and H.B. Menz, “Incorrectly fitted footwear, foot pain and foot disorders: a systematic search and narrative review of the literature”, *Journal of Foot and Ankle Research*, Vol. 11, No. 43, 2018, https://doi.org/10.1186/s13047-018-0284-z.

[2] K. Stanković et al., “Three-dimensional quantitative analysis of healthy foot shape: a proof of concept study”, *Journal of Foot and Ankle Research*, Vol. 11, No. 8, 2018, https://doi.org/10.1186/s13047-018-0251-8.

[3] A.S. Sheikh et al., “A deep learning system for predicting size and fit in fashion e-commerce,” in *RecSys '19: Thirteenth ACM Conference on Recommender Systems*, Copenhagen, 2019, https://doi.org/10.1145/3298689

[4] Y.C. Lee, et al., *“*Comparing 3D foot scanning with conventional measurement methods*”*, *Journal of Foot and Ankle Research* Vol. 7, No. 44, 2014, https://doi.org/10.1186/s13047-014-0044-7

[5] A. Revkov and D. Kanin, in *"FITTIN - Online 3D Shoe Try-on"*, *Proc. of 3DBODY.TECH 2020 – 11th Int. Conf. and Exh. on 3D Body Scanning and Processing Technologies*, Online/Virtual, Nov. 2020, https://doi.org/10.15221/20.58.

[6] D. Omrcen and A. Jurca, "Shoe Size Recommendation System Based on Shoe Inner Dimension Measurement", in *Proc. of 2nd Int. Conf. on 3D Body Scanning Technologies*, Lugano, Switzerland,
2011, https://doi.org/10.15221/11.158.

[7] Küper, K. et al. „Bestimmung von Schuhinnenmaßen—Zerstörungsfreie Werkstoffprüfung mittels Computertomographie“, *German Journal of Foot and Ankle Surgery*, Vol. 3, No. 3, 2005, pp. 159–163, https://doi.org/10.1007/s10302-005-0129-5.

[8] J. Jo, and H. Park, (2020) “Fit of Fire Boots: CT (Computerized Tomography) Scan and 3D Simulation”, in *Proc. of Int. Textile and Apparel Association Ann. Conf.*, Virtual / Online, 2020, https://doi.org/10.31274/itaa.11921.

[9] J. Wittmann et al., Generation of a 3D model of the inside volume of shoes for e-commerce applications using industrial x-ray computed tomography, Engineering Research Express, Vol. 3, No. 4, 2021, https://doi.org/10.1088/2631-8695/ac43c8.

[10] M. Leipert et al., „Three Step Volumetric Segmentation for Automated Shoe Fitting”, in *eJournal of Nondestructive Testing*, Vol. 28, No. 3, 2023, https://doi.org/10.58286/27736.

[11] A. Thompson and R. Leach, “Introdction to Industrial X-Ray Computed Tomography,” in *Industrial X-Ray Computed Tomography*, S. Carmignato et al. Eds. Cham: Springer, 2018, https://doi.org/10.1007/978-3-319-59573-3_1.

[12] P. Hermanek et al., “Principles of X-ray Computed Tomography,” in *Industrial X-Ray Computed Tomography*, S. Carmignato et al. Eds. Cham: Springer, 2018, https://doi.org/10.1007/978-3-319-59573-3_2.

[13] A. Stolfi et al., “Error Sources,” in *Industrial X-Ray Computed Tomography*, S. Carmignato et al. Eds. Cham: Springer, 2018, https://doi.org/10.1007/978-3-319-59573-3_5.

[14] J. Schlemper et al., “Attention gated networks: Learning to leverage salient regions in medical images”, Medical Image Analysis, Vol. 53, 2019, pp. 197-207, https://doi.org/10.1016/j.media.2019.01.012

[15] J.B.T.M. Roerdink, and A. Meijster. “The watershed transform: definitions, algorithms, and parallelization strategies,” in Fundamenta Informaticae 41 (2000), pp. 187–228.

[16] Y. Li et al., “Fully Convolutional Networks for Panoptic Segmentation”, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Nashville, USA, 2021, https://doi.org/10.1109/CVPR46437.2021.00028.

[17] T.S. Newmann, and H. Yi, “A Survey of the Marching Cubes Algorithm,” in Computers & Graphics Vol. 30, No. 5, 2006, pp. 854–879, https://doi.org/10.1016/j.cag.2006.07.021.

[18] L. Feldkamp et al., “Practical Cone-Beam Algorithm,” *Journal of the Optical Society of America* Vol. 1, No. 6, 1984, pp. 612-619, https://doi.org/10.1364/JOSAA.1.000612.

[19] A. Fedorov et al., “3D Slicer as an Image Computing Platform for the Quantitative Imaging Network,” *Magnetic Resonance Imaging*, Vol. 30, No. 9, 2012, pp. 1323-41, https://doi.org/10.1016/j.mri.2012.05.001.

[20] Y. Zhang, „Convolutional Neural Network Based Metal Artifact Reduction in X-Ray Computed Tomography”, *IEEE Transactions on Medical Imaging*, Vol. *37 No.* 6, 2018, pp. 1370–1381. https://doi.org/10.1109/TMI.2018.2823083.

[21] C. Zhang, and Y. Xing, "CT artifact reduction via U-net CNN," in Proc. SPIE Medical Imaging 2018: Image Processing, Houston, Texas, USA, 2018, https://doi.org/10.1117/12.2293903.

[22] L. Gjesteby et al., "Metal Artifact Reduction in CT: Where Are We After Four Decades?," in *IEEE Access*, Vol. 4, pp. 5826-5849, 2016, https://doi.org/10.1109/ACCESS.2016.2608621.