Synthetic Data Generated from CT Scans for Patient Pose Assessment

Manuel Laufer[1] M.LAUFER@UNI-LUEBECK.DE
Dominik Mairhöfer[1] D.MAIRHOEFER@UNI-LUEBECK.DE
Malte Sieren[2] MALTE.SIEREN@UKSH.DE
Hauke Gerdes[2] HAUKE.GERDES@UKSH.DE
Fabio Leal dos Reis[2] FABIO.LEALDOSREIS@UKSH.DE
Arpad Bischof[2,3] ARPAD.BISCHOF@UKSH.DE
Thomas Käster[4] TK@PRCMAIL.DE
Erhardt Barth[1] ERHARDT.BARTH@UNI-LUEBECK.DE
Jörg Barkhausen[2] JOERG.BARKHAUSEN@UKSH.DE
Thomas Martinetz[1] THOMAS.MARTINETZ@UNI-LUEBECK.DE
[1] *Institute for Neuro- and Bioinformatics, University of Lübeck, Germany*
[2] *University Medical Center Schleswig-Holstein, Lübeck, Germany*
[3] *IMAGE Information Systems Europe GmbH, Rostock, Germany*
[4] *Pattern Recognition Company GmbH, Lübeck, Germany*

Editors: Under Review for MIDL 2025

Abstract

An adequate diagnostic quality of radiographs is essential for reliable diagnoses and treatment planning. The patient's pose during radiography is one of the most important factors determining the diagnostic quality. Since patient positioning is difficult and not standardized, an automated AI-based approach using depth images to automatically assess the patient's pose before the radiograph has been taken would be helpful. Due to regulatory hurdles, however, it is difficult in practice to acquire the required depth images and corresponding radiographs. In this paper, we present a framework that can generate such training data synthetically from Computer Tomography scans. We further show that by pretraining on our generated synthetic dataset consisting of 3077 image pairs of upper ankle joints, the pose assessment of real upper ankle joints can be improved by up to 11 percentage points.

Keywords: patient pose assessment, synthetic data generation, diagnostic quality, CT scan, time-of-flight cameras, radiography, deep learning

1. Introduction

The diagnostic quality of radiographs is essential for making reliable diagnoses and planning treatments. Radiographs of inadequate diagnostic quality often lead to retakes and thus to increased radiation exposure for the patient and increased costs for the hospital. In the worst case, inadequate diagnostic quality can lead to incorrect treatment and misdiagnosis. The most important factor affecting the diagnostic quality of a radiograph is the pose of the patient at the time the radiograph is taken (Little et al., 2017). Furthermore, patient positioning is error-prone, as it is not standardized and depends heavily on the patient and the experience of the radiographer, who is also often under time pressure.

To assist the radiographer in positioning the patient and to protect the patient from increased radiation dose, an automatic pose assessment would help. By attaching two Time-of-Flight (ToF) cameras to the X-ray device, we were able to show recently that depth images of anatomical preparations of upper ankle joints contain information that can lead to high accuracy pose assessment (Laufer et al., 2024). In order to determine a correspondence between the depth image of the pose and the diagnostic quality of the radiograph, the radiograph and the depth image must be taken simultaneously and labeled with their diagnostic quality. The depth image and the label can then be used to train neural networks to predict the diagnostic quality of the radiograph before the radiograph is even taken. However, radiographing subjects without an indication is problematic. In particular, intentionally radiographing subjects in non-diagnostic poses, which are necessary for the training, is ethically difficult to justify. Finally, using cameras in live clinical practice is not readily possible for data protection and regulatory reasons. Working with anatomical preparations as a solution is not scalable, and it is arguable whether the movement apparatus of anatomical specimens is identical to that of living subjects.

To address this data challenge, in this paper we present a framework that synthetically generates the required image pairs of depth images and radiographs from Computer Tomography (CT) scans. CT scans that have already been taken can thus be used retrospectively to create a data set of any size, which makes the approach scalable. It is furthermore possible to intentionally generate non-diagnostic poses by selectively adjusting the CT scans. We show that by pretraining on our generated synthetic dataset of upper ankle joints, the pose assessment of real upper ankle joints can be improved by up to 11 percentage points (pp). We will publish the synthetic depth images along with the diagnostic quality labels.

2. Related Work

The generation of synthetic radiographs from CT scans, known as digitally reconstructed radiographs (DRR), is a well-researched field. There are methods based on a forward projection of the CT (Unberath et al., 2018), methods based on a physical simulation (Badal and Badano., 2011) and generative models (Liu and Lin, 2023; Keerthi et al., 2024). Chougule et al. (2013) and Olya Grove and Piegl (2010) present slice-based approaches for generating synthetic point clouds and NURBS from CT scans. To learn multimodal registration with point clouds and CT scans, Saiti and Theoharis (2022) create synthetic point clouds from CT scans using the Marching Cubes Algorithm (MCA)(Lewiner et al., 2012). However, to the best of our knowledge, there is no framework that generates synthetic depth image *and* corresponding radiograph pairs for different view angles from CT scans in order to use them for training patient pose assessment models.

3. Framework

Our framework, which is implemented via Open3d's (Zhou et al., 2018) graphical visualization, is shown in Figure 1 and described in the following sections:

3.1. Preprocessing

In order to create a synthetic depth image of the target anatomy from a CT scan, the target anatomy must first be extracted from the CT. For this purpose, the scan is converted into a point cloud using the Marching Cubes Algorithm (MCA). The relevant parameters are the pixel spacing of the CT and the level threshold value for the MCA, which in this case is defaulted to a Hounsfield (HU) value of -500 so that air around the patient is removed. This conversion from the CT array to the point cloud includes a transformation T_{CT}^{P} from the CT coordinate system CT to the patient coordinate system P. The point cloud is now cropped to the target anatomy to simplify subsequent steps and calculations. Since only the target anatomies surface is relevant for the synthetic depth image, further MCA runs, clustering algorithms such as DBSCAN (Ester et al., 1996) and cropping are applied to remove the imaging table and points inside and outside the surface. In order not to only generate the pose in which the patient was during the CT scan, but also synthetic depth images of other poses, especially non-diagnostic poses, the target anatomy can be brought into other poses by specific rotations of the point cloud. The axis of rotation is strongly dependent on the target anatomy. For the upper ankle joint, it is the longitudinal axis which is positioned in the point cloud such as to mimic human leg rotation. This axis passes through the center of the upper ankle joint. See Figure 3(a) in Section A for illustrations.

3.2. Augmentation

By determining the normal vectors of the point cloud, it is possible to shift the point cloud both in the direction of the normal vector outwards and against the normal vector direction inwards. This allows patients with different shapes to be simulated and the amount of data generated to be increased, see Figure 3(b) in Section A. The label for diagnostic quality applies to all augmentations of a particular pose, as it can be assumed that the anatomy decisive for the quality, in particular the position of the bones in relation to each other, does not change with minor displacements along the normal vectors.

3.3. Scene Composition and Synthetic Depth Image Generation

To generate realistic synthetic depth images, it is beneficial to embed the target anatomy in a realistic scene. This can be achieved, by recording the X-ray room including the imaging table with the depth camera at desired position so that the target anatomy can then be placed on the imaging table under the X-ray device, see Figure 3(c) in Section A. The exact position of the target anatomy is selected so that the X-ray beam of the X-ray device passes through the axis of rotation of the target anatomy. Since a realistic environment and positioning of the target anatomy has been established, it is possible for the user to easily determine the range of rotation of the target anatomy, including intentionally chosen non-diagnostic poses. Both the rotation and the initial embedding of the target anatomy in the X-ray room can be described as a further transformation T_{P}^{ToF} from P to the ToF coordinate system ToF. From the point clouds of the target anatomy and the X-ray room, a 2D projection yields the synthetic depth image by taking into account camera-specific intrinsic and extrinsic parameters and any distortion coefficients. In this way, synthetic depth images can be produced for every angle of rotation as well as for every augmentation. The transformations for each individual pose are used in the next step as described below.

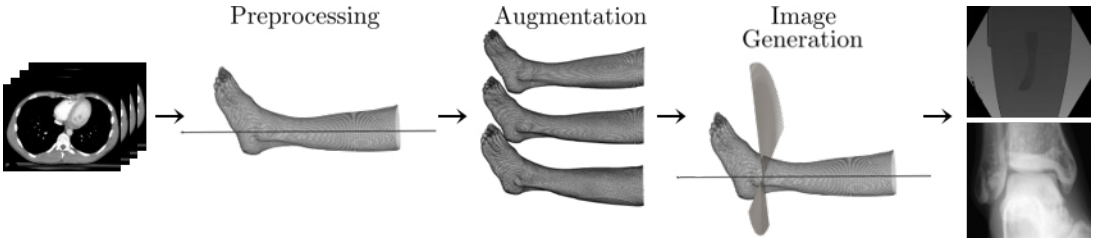

Figure 1: Schematic overview of the framework. The CT scan as input is passed through the steps described in Section 3 to eventually synthetic radiographs and depth images are generated.

3.4. Synthetic Radiograph Generation

When generating synthetic radiographs, it is crucial that all depicted anatomical features are identical with real radiographs in the same position, as otherwise it could lead to false labels of the corresponding depth images. A generation based on a physical model was therefore preferred to methods based on deep learning. Since only a region smaller than the one on the depth images in Section 3.1 must be visible on the radiograph, only the target anatomy is cropped out of the CT in a first step. The cropped CT voxels are then converted into material and mass density voxels. The material voxels each contain one of the materials air, soft tissue, bone or titanium, based on the HU value of the respective voxel. Similarly, the mass density voxels contain the density of the material adjusted by the HU value.

For each pair of material and mass density voxels, multiple radiographs of different positions are generated. Instead of changing the position of the target anatomy, which would need rotation and interpolation of the voxels, the position of the X-ray device in relation to the target anatomy is changed, see Figure 3(*d*) in Section A. The corresponding position of the X-ray device in the CT coordinate system can be achieved by using the inverted transformations $(T_P^{ToF})^{-1}(T_{CT}^{P})^{-1}$. To generate the radiograph the tool MCGPU (Badal and Badano., 2011) was used together with the material properties from PENELOPE 2006 material files (Salvat et al., 2006) to simulate $2 \cdot 10^{10}$ X-ray beam paths. The resulting raw image, containing the energy that reached the detector for each pixel, is then converted to a synthetic radiograph using a non-linear value mapping, to obtain the same look as a real radiograph.

Although the synthetic radiograph is not as detailed as a real radiograph, it is suitable for assessing the diagnostic quality of the pose. Examples of synthetically generated depth images and corresponding radiographs are shown in Figure 4 in Section B

4. Datasets

The two datasets used in this paper consist of depth images from two camera views and corresponding radiographs, which have been assessed regarding their diagnostic quality.

4.1. Synthetic Dataset

Using the framework proposed in Section 3, we were able to generate 3077 image pairs of synthetic radiographs and depth images of upper ankle joints from 10 CTs of different patients. The anonymized CT scans were selected to contain flexed upper ankle positions and exclude clutter such as tubes or screws. From the 10 CTs, a total of 17 upper ankle joints were extracted and rotated medially around the longitudinal axis in an angular range of 90°. A synthetic depth image was generated for each half degree, i.e. a total of 181 poses per foot. This is done for each augmentation and for two camera views, resulting in a total of 18462 depth images. Since the synthetic radiograph image is the same for the camera view and augmentations, one synthetic radiograph was created for each of the 181 poses resulting in a total of 3077 synthetic radiographs. Each of these radiographs was assessed by 4 radiologists according to its diagnostic quality on a scale of 1 to 3 in steps of 0.5. A diagnostic quality of 1 is ideal and a diagnostic quality of 3 is inadequate. The deciding factor in the assessment of the upper ankle joint is the visibility of the joint space, see Figure 4 in Section B. Radiographs with a label in the interval of [1,2.5) can be furthermore classified as *diagnostic* and anything with a label above that as *non-diagnostic*. To the best of our knowledge, this is by far the largest dataset linking depth images to diagnostic quality.

4.2. Anatomic Preparation Dataset

In Laufer et al. (2024) we captured two anatomical preparations - a left and a right lower leg of two women in 174 different poses by two ToF cameras. Parallel to the depth images from two different views, the radiograph of the upper ankle joint were also captured. The subjects were not only rotated medially around the longitudinal axis but also flexed in three different positions of the ankle joint. The radiographs were also evaluated by 4 radiologists in the same manner and on the same scale described in Section 4.1. More detailed information on this published dataset is found in Laufer et al. (2024).

5. Experiments and Results

With the following experiments we intend to answer two questions:

1. Can a neural network be trained on the generated synthetic depth images to assess the pose with high accuracy?

2. If so, can the neural network be finetuned on real data in order to improve the results of the patient's pose assessment?

To answer the first question, a 3-fold cross-validation was performed on the synthetic dataset. The test set, consisting of three upper ankle joints, was randomly selected so that no subject from the test set appears in the training set. To clarify whether the augmentation of the depth images has a benefit, the training was carried out with and once without augmentation. The results are shown in Table 1.

In order to investigate whether the synthetic data are beneficial, we evaluated whether pretraining with the synthetic data improves performance when finetuning on real data.

As the finetuning dataset consisted of only two anatomical preparations, the leave-one-out dataset splitting strategy was used, i.e. training on one preparation and testing on the other, and vice versa. In this experiment, it is furthermore helpful to compare the results with those of a model that was trained without pretraining from scratch on the anatomical preparations dataset and one that was previously pretrained on another dataset in our case on ImageNet (Deng et al., 2009) and then finetuned with the anatomical preparations' dataset. Furthermore, it was evaluated whether camera-specific pretraining (Section 5.1) and the augmentation of the data during pretraining have an influence on the results. The results are shown in Table 2

Table 1: Results of the experiments conducted solely on the synthetic dataset and evaluated by the metrics described in Section 5.2 regarding the impact of augmentation of the synthetic depth images. Note that training with augmented data can improve the results for all metrics.

Metric	without augmentation	with augmentation
MAE	0.23±0.02	**0.21**±0.02
Correlation r_s	0.85±0.03	**0.87**±0.02
Accuracy [%]	85.29±2.34	**87.6**±2.6
Diag. Acc. [%]	85.45±2.77	**86.96**±2.47
Sens. [%]	91.5±1.5	**92.82**±1.8
Spec. [%]	76.9±6.01	**79.0**±5.88

Table 2: Results of the experiments without pretraining, with pretraining on ImageNet and pretraining with the synthetic dataset and subsequent finetuning on the anatomical preparation dataset, using the metrics and methods described in Section 5. Note that pretraining with synthetic data outperforms models without pretraining and with pretraining on ImageNet.

Metric	from scratch	pretrained on ImageNet	pretrained on synthetic dataset			
			unified camera view		camera view specific	
			wo/ aug.	w/ aug.	wo/ aug.	w/ aug.
MAE	0.28±0.04	0.31±0.05	**0.22**±0.04	0.24±0.04	0.23±0.03	**0.22**±0.06
Correlation r_s	0.88±0.04	0.9±0.03	**0.92**±0.03	0.9±0.04	0.9±0.03	0.91±0.04
Accuracy [%]	79.25±6.98	77.37±8.22	89.03±5.51	86.91±6.14	**90.45**±4.73	90.07±7.64
Diag. Acc. [%]	89.08±3.75	84.2±2.43	90.93±3.0	91.91±4.64	**92.8**±2.14	92.43±4.05
Sens. [%]	91.21±4.23	**93.79**±3.19	92.65±7.45	90.9±±0.29	87.64±4.51	91.16±6.72
Spec. [%]	88.66±4.61	80.05±4.6	89.84±5.49	92.24±6.59	**95.34**±3.91	93.1±6.27

5.1. Training

Since in Laufer et al. (2024) the separate network architecture with paired evaluation method performs best, we use it in this paper, to address the multiview problem. In this architecture, there is one network for each different camera view. The outputs of the two networks are then averaged, see Figure 2(a). As models, we used the EfficientNet-B0 (Tan and Le, 2019). Due to the underlying order of the labels, we have modeled the problem as a regression. All experiments were implemented using the PyTorch Lightning framework version 2.3.3 (Falcon and The PyTorch Lightning team, 2019). In order to be comparable to our previous work (Laufer et al., 2024), we have used the same hyperparameters for training, preprocessing of the depth images and augmentation. Note that possible differences in the results in comparison to Laufer et al. (2024) are due to different implementations and updated libraries. The pretraining was identical to the training from scratch and the finetuning. All experiments, including the 3-fold cross validation and leave-one-out strategy were repeated 10 times with different seeds. The results were then averaged.

Since the separate network architecture has one model per camera view, it is possible to pretrain it in two different ways: In the *unified camera view* approach, both models are pretrained on images from both camera perspectives and then only finetuned on images from one camera perspective, see Figure 2(c). This way, all models are pretrained equally with the full number of synthetic depth images. With the *camera view specific* approach, the model that is only finetuned on one camera view is also only pretrained on synthetic depth images from the same camera view, see Figure 2(b). This effectively halves the amount of pretraining data, but the finetuning is more specific.

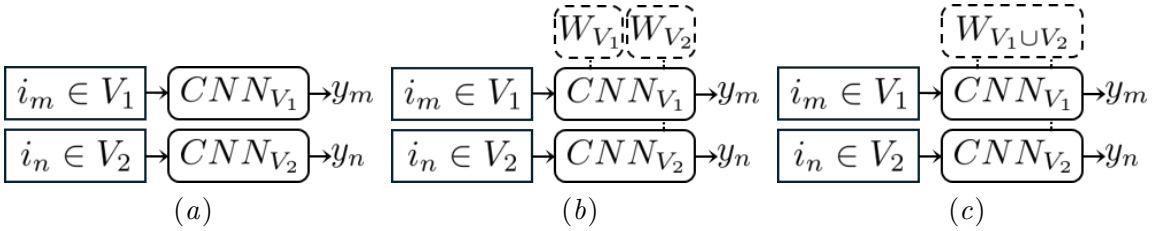

Figure 2: 2(a) shows the separate network architecture: individual images from each view (V_1, V_2) are used to train two separate CNNs. 2(b) shows the *camera view specific* approach, where the CNNs are initialized with the weights (W_{V_1}, W_{V_2}) obtained by pretraining with the corresponding view. 2(c) shows the *unified camera view* approach, where the CNNs are initialized with the weights obtained by pretraining with images from both views.

5.2. Metrics

In addition to the **Mean Absolute Error** (MAE) and the **Spearman correlation** two accuracies are used: the **Accuracy** measures how often the prediction differs less than by 0.5 from the label. The **Diagnostic Accuracy** measures how often the prediction of whether a depth image is diagnostic or non-diagnostic is correct. I.e., whether the label and

prediction both are below or above the 2.5 threshold. Since it is worse to classify an image that is not diagnostic as diagnostic, the **Sensitivity** and **Specificity** are also calculated for the diagnostic accuracy, sensitivity being more important than specificity in this case.

5.3. Results

The high accuracy of 87.6% in the experiments based on the synthetic dataset (Table 1) show that synthetic depth images can be used to learn to assess poses and that training with augmented synthetic depth images results in an improvement in all metrics compared to training without augmentation. The results in Table 2 show that pretraining with the synthetic data improves training on real data. Despite the fact that the sensitivity of the diagnostic quality is highest (93.79%) for the model pretrained on ImageNet, the other metrics show, that there is no benefit, compared to training from scratch. This suggests that the features learned on ImageNet are not useful for the task. This is different for the features learned by pretraining on the synthetic depth images. Except for sensitivity, all metrics for the models pretrained on the synthetic depth images are improved over those without pretraining. The accuracy even increases by about 11 pp to 90.45% for the *camera view specific* approach without augmentation. With the same model, the diagnostic accuracy can also be improved by 3 pp.

Moreover, the *camera view specific* approach is performing slightly better according to the more important metrics accuracy metrics, which is presumably due to the higher similarity of the pretraining and finetuning dataset. In contrast to the results from Table 1, no advantage can be determined from the use of augmentation of the synthetic depth images in these experiments. This is possibly due to the lack of variance in the shape and the small number of the anatomical preparations, both of which come from relatively thin subjects.

6. Conclusion and Outlook

In this paper, we presented a framework for generating both synthetic depth images and radiographs from CT scans. We have shown that by pretraining on such a synthetic dataset relevant features can be learned, which are useful for the assessment of patients‘ poses on real data. This makes it easier to determine whether the patient's pose would lead to a radiograph with inadequate diagnostic quality before it is taken and thus protect the patient from unnecessary radiation due to a retake.

The main advantage of this framework is that the data acquisition problem, which often exists in the medical context, can be solved by using already available CT scans. It is also possible to adapt the synthetic training data to different X-ray rooms and different ToF cameras in order to generate realistic case-specific training data that is as close to real data as possible. Furthermore, it is possible to investigate which camera positions and how many cameras are best suited for the pose assessment task. Although we have here only shown this for upper ankle joints, we believe that the framework can also be applied to other anatomies.

Ideally our novel way of generating radiographs and depth images can be used beyond the here demonstrated pose-assessment application.

References

Andreu Badal and Aldo Badano. Chapter 50 - fast simulation of radiographic images using a monte carlo x-ray transport algorithm implemented in CUDA. In Wen-mei W. Hwu, editor, *GPU Computing Gems Emerald Edition*, Applications of GPU Computing Series, pages 813–829. Morgan Kaufmann, Boston, 2011. ISBN 978-0-12-384988-5. doi: 10.1016/B978-0-12-384988-5.00050-4.

Vikas Chougule, Arati Mulay, and B. Ahuja. Conversions of ct scan images into 3d point cloud data for the development of 3d solid model using b-rep scheme. 12 2013.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255, June 2009. doi: 10.1109/CVPR.2009.5206848.

Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, page 226–231. AAAI Press, 1996.

William Falcon and The PyTorch Lightning team. PyTorch Lightning, March 2019. URL https://github.com/Lightning-AI/lightning.

R Keerthi, Kuval Kiran, Ss Kiran, and P Likitha. Advancing medical imaging: A comparative exploration of generative adversarial networks for chest x-ray synthesis. In *2024 IEEE International Conference on Computer Vision and Machine Intelligence (CVMI)*, pages 1–7, 2024. doi: 10.1109/CVMI61877.2024.10782852.

Manuel Laufer, Dominik Mairhöfer, Malte Sieren, Hauke Gerdes, Fabio Leal dos Reis, Arpad Bischof, Thomas Käster, Erhardt Barth, Jörg Barkhausen, and Thomas Martinetz. Patient pose assessment in radiography using time-of-flight cameras. In Olivier Colliot and Jhimli Mitra, editors, *Medical Imaging 2024: Image Processing*, volume 12926, page 129261J. International Society for Optics and Photonics, SPIE, 2024. doi: 10.1117/12.3000370. URL https://doi.org/10.1117/12.3000370.

Thomas Lewiner, Hélio Lopes, Antonio Vieira, and Geovan Tavares. Efficient implementation of marching cubes' cases with topological guarantees. *Journal of Graphics Tools*, 8, 04 2012. doi: 10.1080/10867651.2003.10487582.

Kevin J. Little, Ingrid Reiser, Lili Liu, Tiffany Kinsey, Adrian A. Sánchez, Kateland Haas, Florence Mallory, Carmen Froman, and Zheng Feng Lu. Unified Database for Rejected Image Analysis Across Multiple Vendors in Radiography. *Journal of the American College of Radiology*, 14(2):208–216, February 2017. ISSN 1546-1440. doi: 10.1016/j.jacr.2016.07.011.

Jian Liu and Tim H. Lin. A framework for the synthesis of x-ray security inspection images based on generative adversarial networks. *IEEE Access*, 11:63751–63760, 2023. doi: 10.1109/ACCESS.2023.3288087.

Khairan Rajab Olya Grove and Les A. Piegl. From ct to nurbs: Contour fitting with b-spline curves. *Computer-Aided Design and Applications*, 7(sup1):1–19, 2010. doi: 10.1080/16864360.2010.10738807. URL https://doi.org/10.1080/16864360.2010.10738807.

E. Saiti and T. Theoharis. Multimodal registration across 3d point clouds and ct-volumes. *Comput. Graph.*, 106(C):259–266, August 2022. ISSN 0097-8493. doi: 10.1016/j.cag.2022.06.012. URL https://doi.org/10.1016/j.cag.2022.06.012.

Francesc Salvat, José M Fernández-Varea, and Josep Sempau. Penelope-2006: A code system for monte carlo simulation of electron and photon transport. In *Workshop Proceedings*, volume 4, page 7, 2006.

Mingxing Tan and Quoc Le. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. In *International Conference on Machine Learning*, pages 6105–6114. PMLR, May 2019.

Mathias Unberath, Jan-Nico Zaech, Sing Chun Lee, Bastian Bier, Javad Fotouhi, Mehran Armand, and Nassir Navab. DeepDRR – A Catalyst for Machine Learning in Fluoroscopy-Guided Procedures. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 98–106, Cham, 2018. Springer International Publishing. ISBN 978-3-030-00937-3. doi: 10.1007/978-3-030-00937-3_12.

Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018.

Appendix A. Framework Illustrations

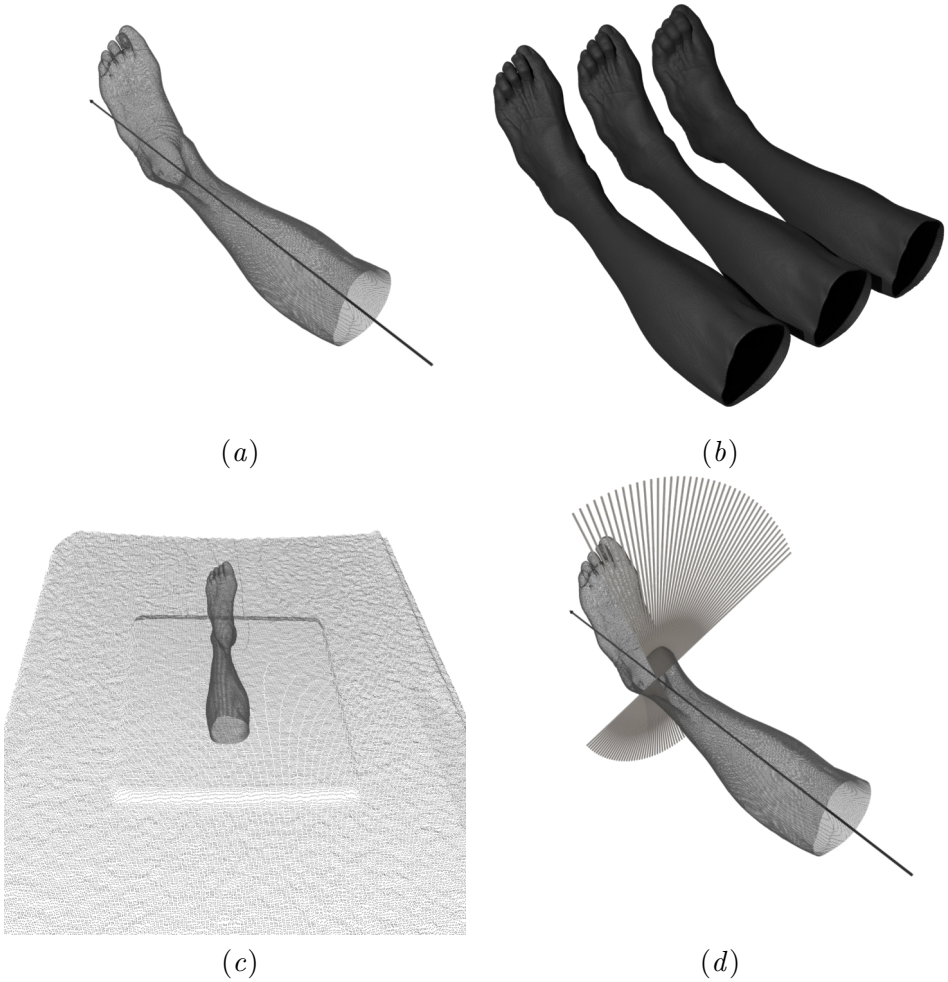

Figure 3: This figure illustrates the different steps described in Section 3. Figure 3(a) shows the target anatomy cut out of the CT scan as a point cloud and the rotation axis, which passes through the center of the upper ankle joint. Figure 3(b) shows the augmentation of the point clouds by moving the points along the direction of the normal vectors. Figure 3(c) shows the target anatomy combined with the previously acquired X-ray room including the imaging table and detector. Figure 3(d) sketches the positions of the X-ray device, which change due to the medial rotation around the longitudinal axis of rotation.

Appendix B. Synthetic Example Images

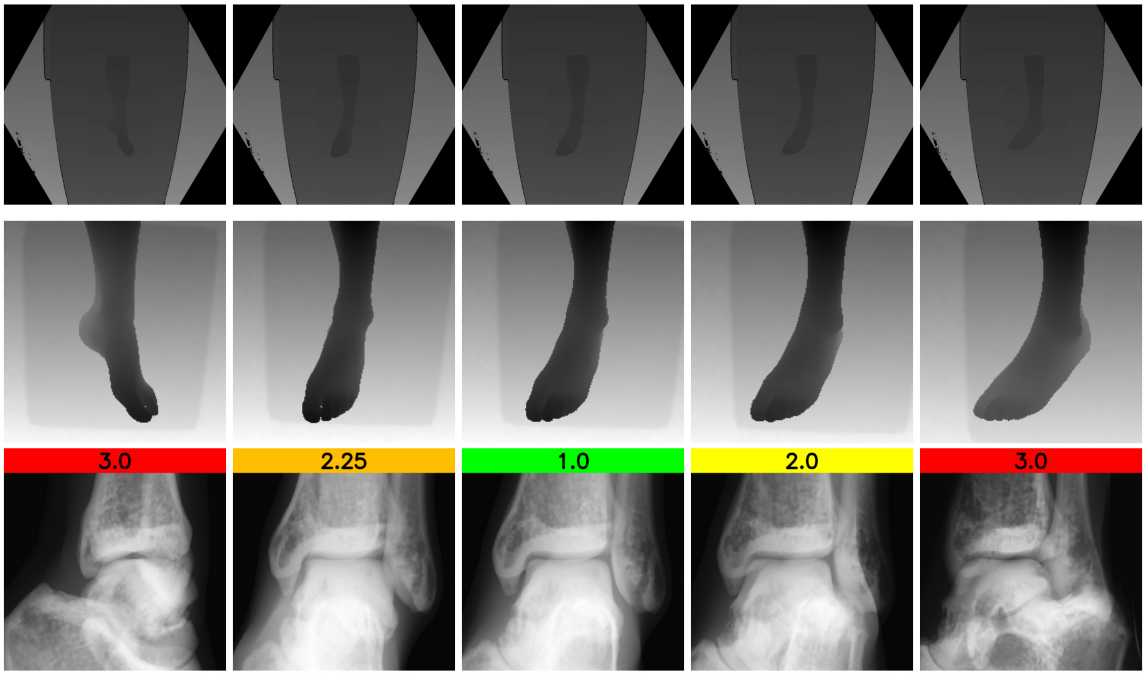

Figure 4: This figure shows the synthetic images generated by using the framework. The first row shows the synthetic depth images that were generated with different rotations of the target anatomy. The second row shows the manually created ROIs for the synthetic depth images from the first row, which are then used for training. The third row shows the synthetic radiographs corresponding to the synthetic depth images, including their diagnostic quality, which has been assessed by the radiologists. Note that only small rotations are necessary to change a diagnostic quality of 1 to a diagnostic quality of 2, which is reflected in the visibility of the joint space in the different images.