3D Segmentation of Hard and Soft Tissue to Facilitate X-ray Image Formation Using Deep Learning

Students: Sean S. Darcy, Zhiyuan Ding, Qiyuan Wu
Mentors: Benjamin Killeen, Mathias Unberath
3/1/2022

Project Background

Fluoroscopy-guided interventions are minimally-invasive procedures during which the surgeon is guided by a continuous stream of X-ray images displayed on a monitor. This class of intervention has become popular in orthopedic surgeries, liver biopsies, pacemaker implantations, catheter insertions, angiography, and a variety of other procedures. Fluoroscopy-guided intervention is a rapidly growing discipline and one that is prime for automation- particularly using deep learning approaches with models trained on intraoperative image data.

Unfortunately, one big problem limiting the success of these automation approaches is the availability of training data. High-quality, expert-labeled ground truth X-ray images are scarce. On top of this, intraoperative images- which would depict tools and image features unique to the operating room environment- are typically discarded and even less available for use in learning. Successful approaches to automation of fluoroscopy-guided interventions require substantial amounts of high quality training data, but this data is difficult to obtain. In addition, labeling this data would be a massive challenge due to the sheer quantity involved. One potential solution to this problem is simulating these X-ray images.

A digitally reconstructed radiograph (DRR) is a simulated X-ray image. A DRR is generated by ‘imaging’ a computed tomography (CT) volume in silico. Annotation and augmentation can be performed on the CT volume as opposed to individual images- reducing workload and promoting valid image characteristics. Until recently, these simulated X-ray images were not very realistic and failed to translate to the clinic. Enter DeepDRR.

DeepDRR was developed here at Hopkins and provides state of the art tools to generate realistic radiographs at training set scale. It renders the most realistic DRRs to date as demonstrated by a pelvis landmark detection task, where DeepDRR radiographs substantially outperformed other DRRs. Despite being more realistic, current DeepDRR images are still not indistinguishable from true radiographs. Displayed in figure 1 is the DRR pipeline, which begins with a segmentation of the CT into 3 components- bone, soft tissue, and air. In reality, there are more than just 3 classes of material in the human body, and each has its own unique properties with respect to interaction with X-rays. To render more realistic DRRs, we should accurately represent these materials and their properties in our in silico projections. Thus, our project focuses on the

segmentation component of the DeepDRR framework.

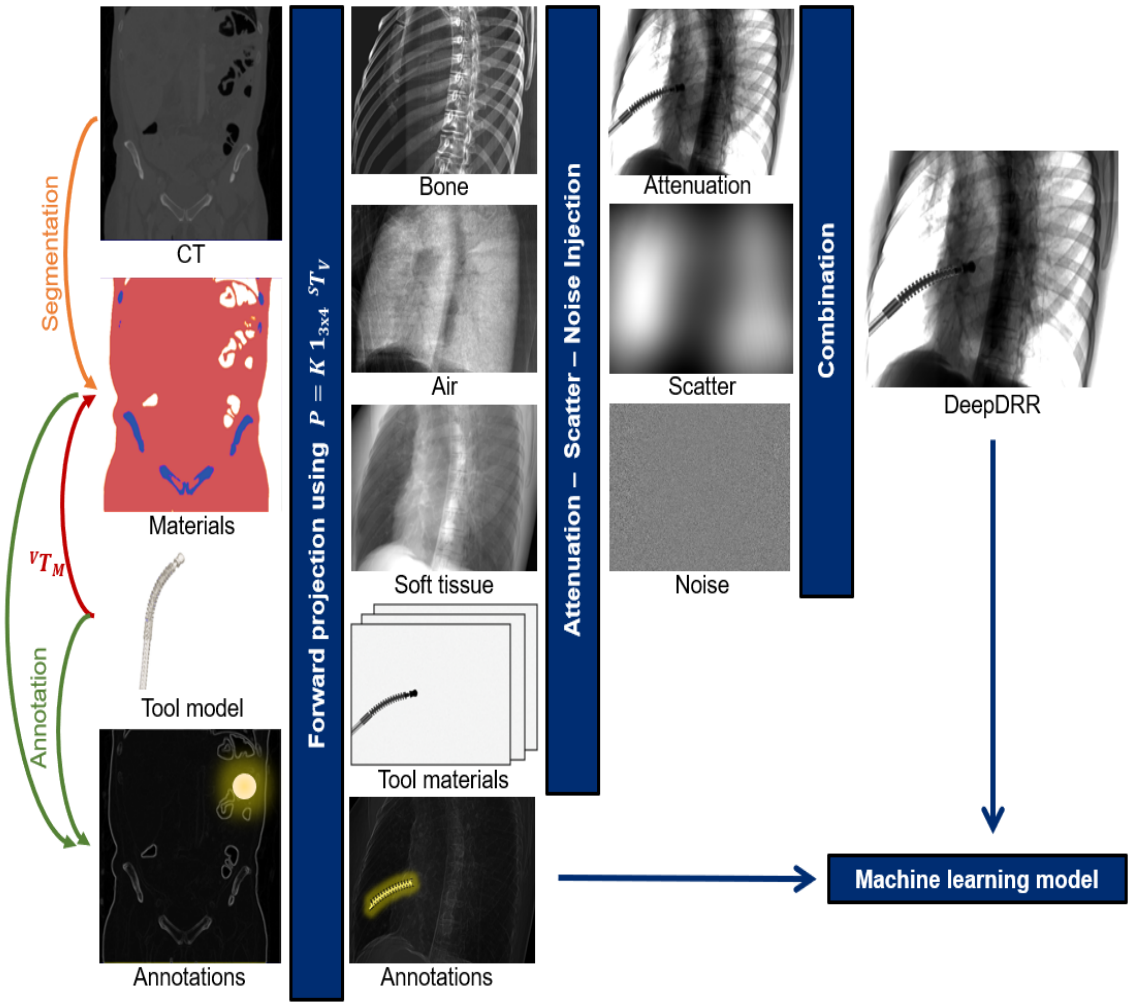


Figure 1. DeepDRR

We want to start by building and testing algorithms that automatically segment tissues of varying absorption rates from CT data. Next, we plan to compare the performance of three distinct 3D segmentation model architectures on at least 2 tissue types and finalize a novel 3D segmentation pipeline. We are prioritizing the cardiac, lung and liver tissues as these were most relevant to fluoroscopy procedures according to our review. We will then integrate our novel 3D segmentation pipeline with the current DeepDRR system with the ultimate goal of improving DRR quality and effectiveness in learning. The final component of our project is to compare the performance of deep learning fluoroscopy and radiology-oriented models trained on either real X-ray images, DeepDRRs with our novel 3D segmentation module, and DeepDRRs as they are rendered today. Specifically, we would test these models on real X-ray images to validate that our improved DRRs effectively train models meant to operate on real, non-simulated data generated in the clinic.

Technical Approach

3D Segmentation of Hard and Soft Tissue

In order to improve the performance of DeepDRR, we need to generate accurate masks for comprehensive classes of organs with varying X-ray interaction properties. Thus, the initial component of our project involves 3D segmentation of full-body computed tomography (CT) data.

Based on our initial literature review of 3D CT segmentation, previous research can be grouped into two categories: end-to-end models and coarse-to-fine frameworks. The former involves building a large network to complete the entire task, such as a U-net with attention strategy. The second strategy follows a coarse to fine process- a multi-model, multi-stage framework for segmentation that attempts to improve segmentation performance after each stage.

The first stage is the 'coarse' segmentation, where an initial segmentation or localization is generated with down-sampled images, as the original ones are often too large to obtain quality segmentations. After this stage, a better localization of the segment target or 'coarse mask' for image cropping is generated. Our 3D segmentation pipeline- currently focused on lung, liver, and cardiac tissue- will generate a coarse mask output. We plan to use existing network structures like the original U-Net, SE-Net (a modification of U-Net with attention), and nnU-Net for this part of the task. Most of these have existing code available online. We will decide on which architecture to implement based on the coarse segmentation results observed as well as computational burden.

The second stage 'fine' segmentation, where partial patches with high resolution from the first stage outputs are sampled for downstream model training. These models execute downstream tasks including small object (e.g. vertebrae) segmentation or texture segmentation for soft tissue. Another important issue here is the sampling strategy. We plan to deploy existing methods such as direct random sampling or coarse mask guidance, although we may end up designing our own patch sampling strategy and checking its performance if necessary.

In the third stage, the previously generated patch results are aggregated. In this stage, we utilize a voting strategy to decide the prediction probability for each voxel and then use post-processing methods such as region connection and edge level sets to generate the final output. Another important point is the assessment of this stage. As our approach integrates multiple models of varying specialties into one comprehensive 3D segmentation framework, we must validate that the segmentations generated are accurate. Thus we will apply a barrage of common segmentation and model evaluation metrics like the Dice coefficient.

Integration with DeepDRR

The second major component of our project involves integrating our novel 3D segmentation pipeline into the DeepDRR framework in order to improve the effectiveness of using DRRs in learning applications. This involves computing the contribution of each segmented material to the total attenuation density at detector position *u* using the geometry defined by projection matrix $P \in R^{3\times4}$ and X-ray spectral density $p_0(E)$ via ray-tracing.

$$p(\boldsymbol{u}) = \int p(E, \boldsymbol{u}) \mathrm{d}E$$
$$= \int p_0(E) \exp\left(\sum_{m \in M} \delta\left(m, M(\boldsymbol{x})\right) (\mu/\rho)_m(E) \int \rho(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{l}_{\boldsymbol{u}} \right) \mathrm{d}E \,, \quad (1)$$

where $\delta(\cdot,\cdot)$ is the Kronecker delta, $\mathbf{l_u}$ is the 3D ray connecting the source position and 3D location of detector pixel *u* determined by P, $(\mu/\rho)_m(E)$ is the material and energy dependent linear attenuation coefficient, and $\rho(x)$ is the material density at position x derived from HU values. Our main task here will be to obtain the linear attenuation coefficient for each of of segmentation masks (m) corresponding to various tissue types. The projection domain image p(u) is then used as input to DeepDRR's scatter prediction ConvNet.

Validation

Validation of our proposed solution involves two components. First, we need to assess the performance of our overall 3D segmentation pipeline to ensure that the segmentation outputs are accurate for each tissue. Second, we must demonstrate that our novel 3D CT segmentation pipeline achieves our ultimate goal of improving DRR quality and effectiveness in model training.

Validation of 3D Segmentation Results

Segmentation model performance can be measured via comparison of output masks with ground truth annotation. We will deploy standard metrics such as the Dice coefficient- which measures a normalized overlap rate of two distributions- to compare output masks to ground truth segmentations. We will collect these metrics for segmentations of each tissue type for each model architecture applied on a standard dataset. These metrics will enable us to piece together our 3D segmentation module as a "mosaic of models," which assembles the highest performing models for each tissue type and outputs the most accurate segmentation masks for the tissues we need in a computationally efficient manner.

Validation of DRR Improvement

The ultimate goal of integrating a novel 3D segmentation pipeline with DeepDRR is to improve DRR effectiveness in model training. In order to assess this improved DRR quality, we plan to compare the performance of downstream-task models trained on either real X-ray images, DeepDRRs generated with our novel 3D segmentation module, or DeepDRRs as they are rendered today. These models must learn from and take X-ray images as input in order to achieve some downstream task. For example, there are models that detect, localize, or even segment key landmarks on bone structures. We are strategically selecting downstream radiology or fluoroscopy learning tasks covering the variety of tissues we segment, in addition to considering clinical relevance and availability of data.

After separately training three instances of the same downstream-task model with either real X-ray images, DeepDRRs generated with our novel 3D segmentation module, or DeepDRRs as they are rendered today, we would test these models on *only* real X-ray images to validate that our improved DRRs effectively train models meant to operate on real, non-simulated data generated in the clinic. We expect to see improved testing accuracy for models trained on DeepDRRs generated with our novel 3D segmentation module compared to models trained on DeepDRRs as they are rendered today, as the simulated training set of the former is a better representation of real X-ray images for a model to learn from than the latter. Our hope is that the testing accuracy of the model trained on DeepDRRs generated with our novel 3D segmentation module gets as close as possible or even exceeds that of the model trained on real X-ray images with the potential to learn more from a larger simulated dataset.

Deliverables

Minimum:

Scripts of a validated segmentation pipeline for bone, soft tissue, and air in full-body CT.
Comparison report of existing models on different type of tissues.

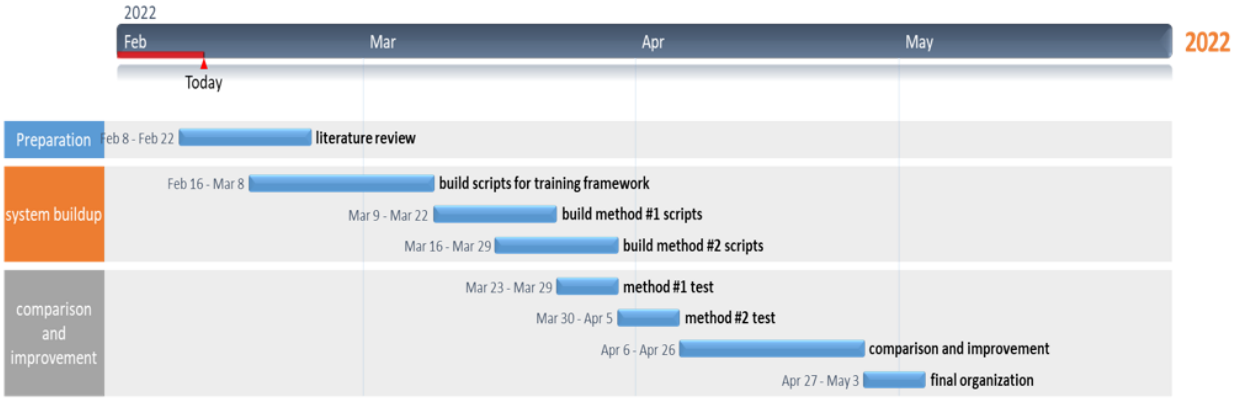
Expected: Minimum +

Scripts and results for segmenting lung, cardiac, liver in full-body CT.
Integrated scripts into the DeepDRR simulation.

Maximum: Expected +

Validation scripts and results on full and partial CT with possible data corruptions, fractures, wounds, etc.
Analysis report of the DRR performance (comparison of downstream task models trained on different data types) using the improved segmentation pipeline.

Timeline



Milestones and Dependencies

Milestones	Expected Date
Comparison report of existing models	Feb 22
Scripts for training framework	March 8
Scripts for bone, soft tissue and air segmentation	March 22
Scripts for lung, cardiac and/or liver segmentation	March 27
Report for segmentation results	April 6
Integration to deepDRR system	April 6
Downstream-task model comparison report	April 12

Dependencies	Needed by	Status / Contact	Remedies
Access to the workstation for training models in Dr. Unberath's lab	Feb 28th	**Completed** **Benjamin Killeen**	Run the training process on PC or Google Colab.
Full body CT dataset	Feb 28th	**Completed**	Use partial CT dataset for bones, soft tissue or other tissues.
Pretrained model for specific tasks	March 15th	**Completed**	Drop the optional task.
Desktops / Laptops	Feb 28th	**Completed**	-

Project Management

1. Weekly meetings with mentor (Benjamin Killeen) at Thursday 2:50pm.
2. Additional meetings inside the group (flexible).
3. Group slack channel for communication.
4. Git repository for code management
5. Shared Zotero library for literature management.

Reading List/Sources

Unberath, M., Zaech, J. N., Lee, S. C., Bier, B., Fotouhi, J., Armand, M., & Navab, N. (2018, September). Deepdrr–a catalyst for machine learning in fluoroscopy-guided procedures. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp. 98-106). Springer, Cham.

Liu, P., Han, H., Du, Y., Zhu, H., Li, Y., Gu, F., … & Zhou, S. K. (2021). Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. International Journal of Computer Assisted Radiology and Surgery, 16(5), 749-756.

Payer, C., Stern, D., Bischof, H., & Urschler, M. (2020, February). Coarse to Fine Vertebrae Localization and Segmentation with SpatialConfiguration-Net and U-Net. In VISIGRAPP (5: VISAPP) (pp. 124-133).

Isensee, F., Jaeger, P.F., Kohl, S.A.A. et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat Methods 18, 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z

Tang Y, Gao R, Lee H H, et al. High-resolution 3D abdominal segmentation with random patch network fusion[J]. Medical Image Analysis, 2021, 69: 101894.

Zhang Y, Wu J, Liu Y, et al. A deep learning framework for pancreas segmentation with multi-atlas registration and 3D level-set[J]. Medical Image Analysis, 2021, 68: 101884.

Zhang Y, Li H, Du J, et al. 3D multi-attention guided multi-task learning network for automatic gastric tumor segmentation and lymph node classification[J]. IEEE Transactions on Medical Imaging, 2021, 40(6): 1618-1631.

Schlemper J, Oktay O, Schaap M, et al. Attention gated networks: Learning to leverage salient regions in medical images[J]. Medical image analysis, 2019, 53: 197-207.

Xu X, Lian C, Wang S, et al. Asymmetric multi-task attention network for prostate bed segmentation in computed tomography images[J]. Medical Image Analysis, 2021, 72: 102116.

Nadeem S A, Hoffman E A, Sieren J C, et al. A CT-based automated algorithm for airway segmentation using freeze-and-grow propagation and deep learning[J]. IEEE transactions on medical imaging, 2020, 40(1): 405-418.

Milletari F, Navab N, Ahmadi S A. V-net: Fully convolutional neural networks for volumetric medical image segmentation[C]//2016 fourth international conference on 3D vision (3DV). IEEE, 2016: 565-571.

Hubbell, J.H., Seltzer, S.M.: Tables of X-ray mass attenuation coefficients and mass energy-absorption coefficients 1 keV to 20 MeV for elements Z= 1 to 92 and 48 additional substances of dosimetric interest. Technical report, National Inst. of Standards and Technology (1995)