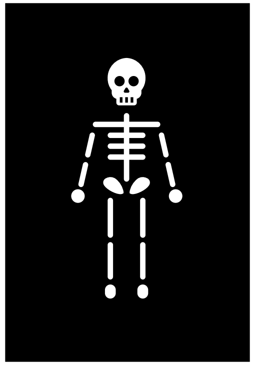
Background Reading Presentation:

3D Segmentation of Hard and Soft Tissue for Simulating X-ray Image Formation with Deep Learning

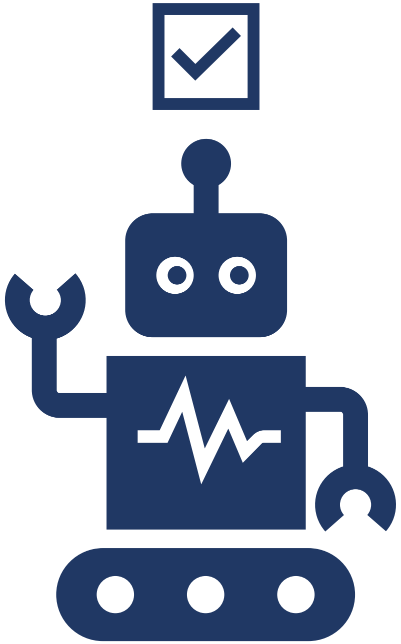
Students: Sean S. Darcy, Qiyuan Wu, Zhiyuan Ding

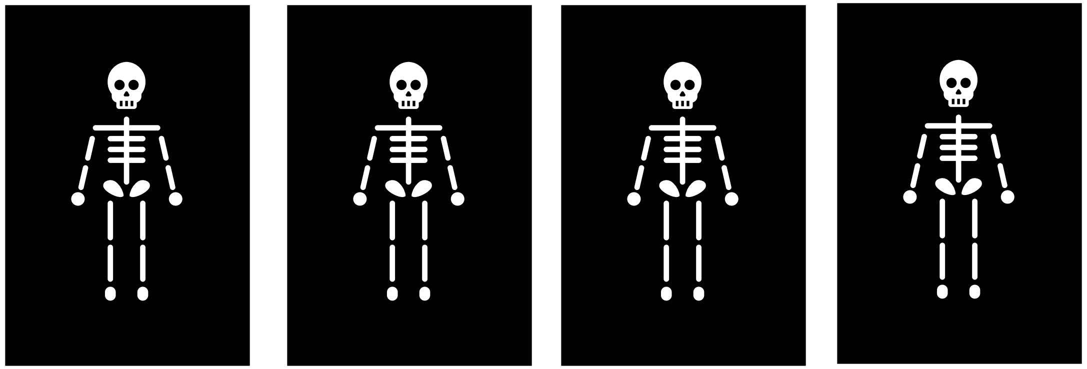
Mentors: Benjamin Killeen, Mathias Unberath

Project Review



Real Clinical Data





Simulated Training Data

In silico simulation of training examples may improve A.I. performance. The goal of this project is to improve the quality of simulated X-ray data.

Background reading

- Papers to be discussed:
  [1] Isensee F, Petersen J, Klein A, et al. nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation. *arXiv:180910486 [cs]*. Published online September 27, 2018.
- [2] Unberath M, Zaech J N, Lee S C, et al. Deepdrr–a catalyst for machine learning in fluoroscopy-guided procedures[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018: 98-106.
- [3] Manuel Schultheiss et al. "Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN based diagnosis to human performance". In: *Scientific Reports* 11.1 (2021), pp. 1–10.

Paper #1 – nnU-Net (Isensee F. et al., 2018)

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 27 Sep 2018]

nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation

Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F. Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, Klaus H. Maier-Hein

Reason for selection

- Performance of 3D segmentation is crucial for DRR
- This paper includes detailed technical approach of nnUNet, which was proved to have state-of-the-art performance on multiple tasks.

nnU-Net (Isensee F. et al., 2018)

Quick Abstract

- Introduces the *self-adapting* framework of nnU-Net ("nonew-UNet")
- Reports nnU-Net performance on various tasks

nnU-Net (Isensee F. et al., 2018)
Methods

- 2D U-net
- 3D U-net
- U-net cascade
- Ensembled inference

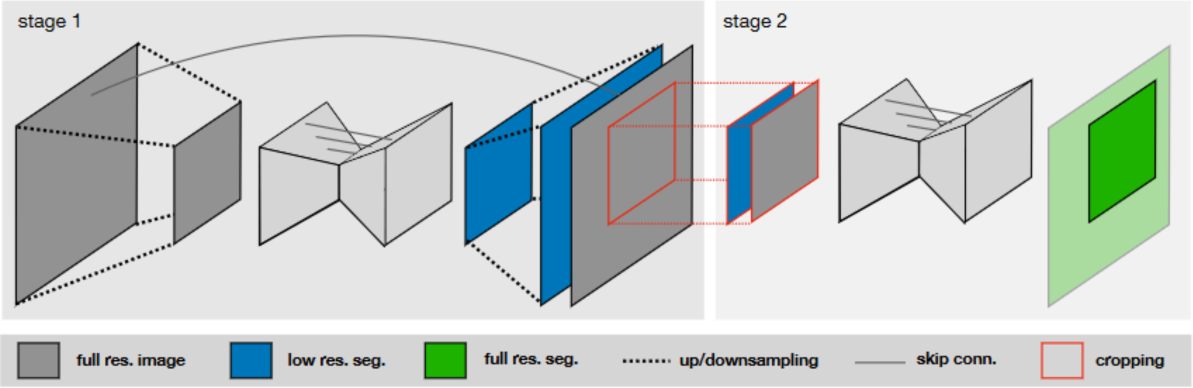


Fig. 1. U-Net Cascade (on applicable datasets only). Stage 1 (left): a 3D U-Net processes downsampled data, the resulting segmentation maps are upsampled to the original resolution. Stage 2 (right): these segmentations are concatenated as one-hot encodings to the full resolution data and refined by a second 3D U-Net.

		2D U-Net	3D U-Net	3D U-Net lowres
BrainTumour	median patient shape	169x138	138x169x138	-
	input patch size	192x160	128x128x128	-
	batch size	89	2	-
	num pool per axis	5, 5	5, 5, 5	-
Heart	median patient shape	320x232	115x320x232	58x160x116
	input patch size	320x256	80x192x128	64x160x128
	batch size	33	2	2
	num pool per axis	6, 6	4, 5, 5	4, 5, 5
Liver	median patient shape	512x512	482x512x512	121x128x128
	input patch size	512x512	128x128x128	128x128x128
	batch size	10	2	2
	num pool per axis	6, 6	5, 5, 5	5, 5, 5
Hippocampus	median patient shape	50x35	36x50x35	-
	input patch size	56x40	40x56x40	-
	batch size	366	9	-
	num pool per axis	3, 3	3, 3, 3	-
Prostate	median patient shape	320x319	20x320x319	-
	input patch size	320x320	20x192x192	-
	batch size	26	4	-
	num pool per axis	6, 6	2, 5, 5	-
Lung	median patient shape	512x512	252x512x512	126x256x256
	input patch size	512x512	112x128x128	112x128x128
	batch size	10	2	2
	num pool per axis	6, 6	4, 5, 5	4, 5, 5
Pancreas	median patient shape	512x512	96x512x512	96x256x256
	input patch size	512x512	96x160x128	96x160x128
	batch size	10	2	2
	num pool per axis	6, 6	4, 5, 5	4, 5, 5

Table 1. Network topologies as automatically generated for the seven phase 1 tasks of the Medical Segmentation Decathlon challenge. 3D U-Net lowres refers to the first stage of the U-Net Cascade. The configuration of the second stage of the U-Net Cascade is identical to the 3D U-Net.

nnU-Net (Isensee F. et al., 2018)

Results

label	BrainTumour 1	BrainTumour 2	BrainTumour 3	Heart 1	Liver 1	Liver 2	Hippoc. 1	Hippoc. 2	Prostate 1	Prostate 2	Lung 1	Pancreas 1	Pancreas 2
2D U-Net	78.60	58.65	77.42	91.36	94.37	53.94	88.52	86.70	61.98	84.31	52.68	74.70	35.41
3D U-Net	**80.71**	**62.22**	**79.07**	92.45	94.11	61.74	**89.87**	**88.20**	60.77	83.73	55.87	77.69	42.69
3D U-Net stage1 only (U-Net Cascade)	-	-	-	90.63	94.69	47.01	-	-	-	-	65.33	79.45	49.65
3D U-Net (U-Net Cascade)	-	-	-	92.40	95.38	58.49	-	-	-	-	**66.85**	**79.30**	**52.12**
ensemble 2D U-Net+ 3D U-Net	80.79	61.72	79.16	**92.70**	94.30	60.24	89.78	88.09	**63.78**	**85.31**	55.96	78.26	40.46
ensemble 2D U-Net+ 3D U-Net (U-Net Cascade)	-	-	-	92.64	95.31	60.09	-	-	-	-	61.18	78.79	45.46
ensemble 3D U-Net+ 3D U-Net (U-Net Cascade)	-	-	-	92.63	**95.43**	**61.82**	-	-	-	-	65.16	79.70	49.14
test set	67.71	**47.73**	**68.16**	**92.77**	**95.24**	**73.71**	**90.37**	**88.95**	**75.81**	**89.59**	**69.20**	**79.53**	**52.27**

Table 2. Mean dice scores for the proposed models in all phase 1 tasks. All experiments were run as five-fold cross-validation. The models that we used for generating our test set submission are highlighted in bold. The dice scores of the test sets are shown at the bottom of the table. Test dice scores in bold denote that at the time of manuscript submission these scores were the highest in the online leaderboard of the challenge (decathlon.grand-challenge.org/evaluation/results).

nnU-Net (Isensee F. et al., 2018)
Pros

- 1. This paper presents an adaptable framework for medical image segmentation, which is *generalizable* for different organ and tissue types.

- 2. This paper provides automated preprocessing and postprocessing, along with many pretrained models, which can be directly used on the dataset we are interested in.

nnU-Net (Isensee F. et al., 2018)
Cons

- 1. The performance on lung segmentation isn't promising.

- 2. Models for some organs are missing or trained on a different modality (e.g. bone, heart). We have to skip these organs when applied to DRR.

nnU-Net (Isensee F. et al., 2018)
Takeaway

- 1. We could use different models for the segmentation of different organs since each model has its best performance on different tasks.

Paper #2 – DeepDRR (Unberath M et al., 2018)

DeepDRR – A Catalyst for Machine Learning in Fluoroscopy-Guided Procedures

Authors: Mathias Unberath, Jan-Nico Zaech, Sing Chun Lee, Bastian Bier, Javad Fotouhi, Mehran Armand, Nassir Navab

Published in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2018

Publisher: Springer International Publishing

Reason for selection

- Main process for CT-based X-ray image simulation
- This paper is the main framework we work on for simulation

DeepDRR (Unberath M et al., 2018)

Quick Abstract

- A CT based X-ray Image Simulation framework was proposed

Key question:

- What components are included in this simulation
- How to realize each sub-process for X-ray image generation

DeepDRR (Unberath M et al., 2018)

Methods

1 Material decomposition

2 Analytic Primary Computation

3 Learning-based Scatter Estimation

4 Noise Injection

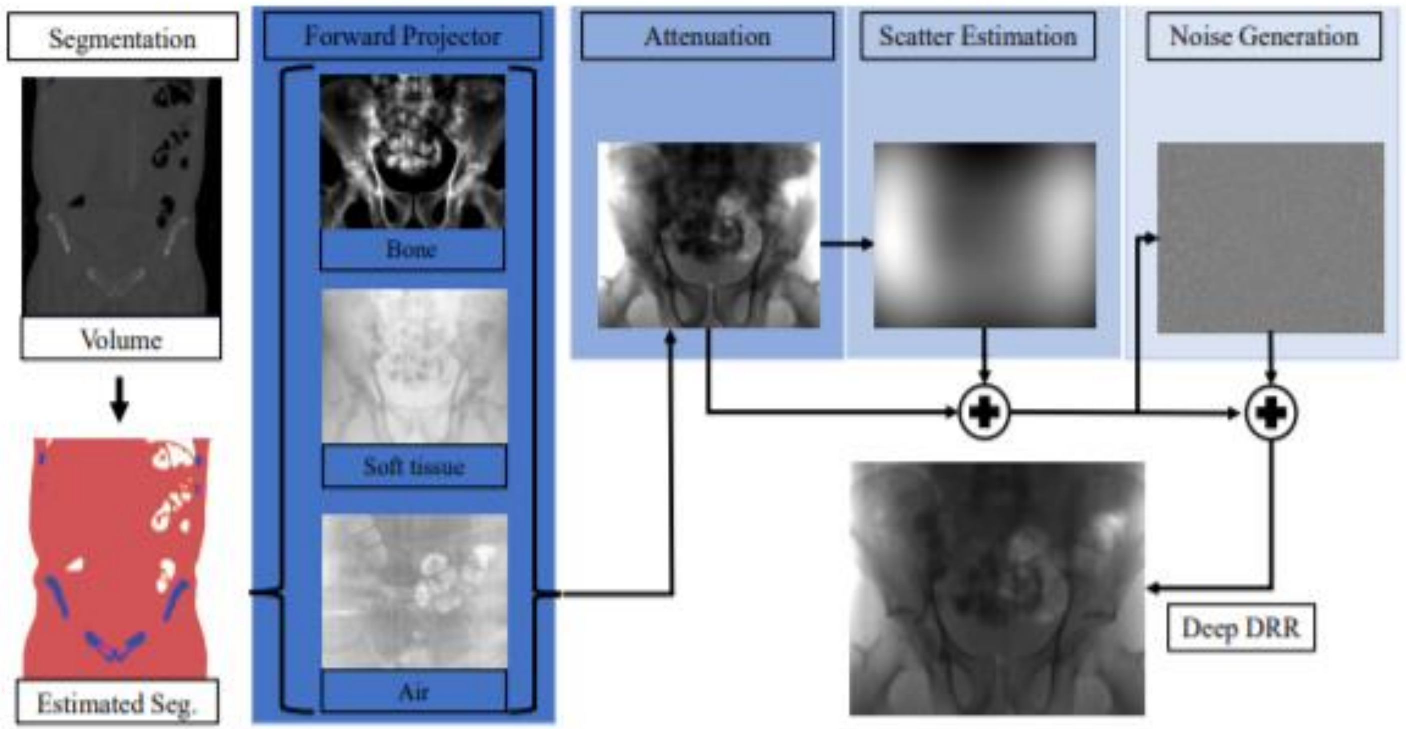


Fig. 1. Schematic overview of DeepDRR.

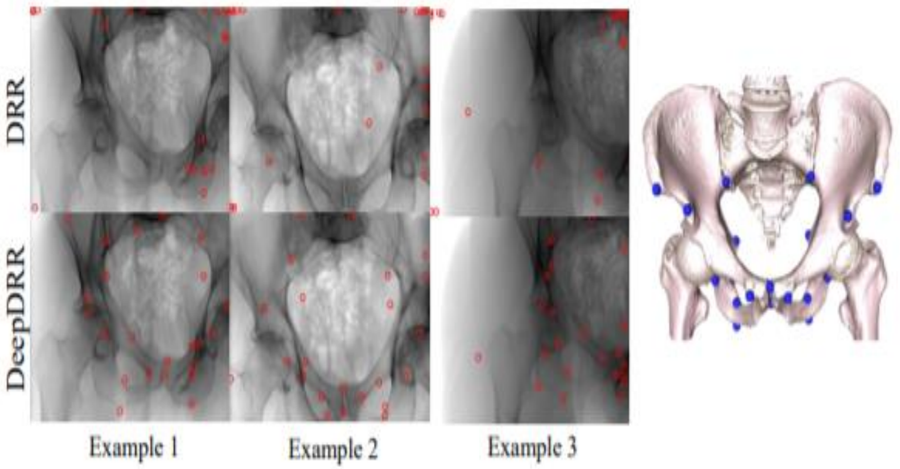


Fig. 3. Anatomical landmark detection on real data of cadaveric specimen using the method detailed in [9]. Top row: Detection results of a model trained on conventional DRRs. Bottom row: Detections of a model trained on the proposed DeepDRRs. No domain adaption or re-training was performed. Right-most image: Schematic illustration of desired landmark locations shown on a training set sample.

DeepDRR (Unberath M et al., 2018)

Pros

1 DeepDRR provides a general framework for CT based X-ray image simulation with useful toolboxes for simulation with tunable parameters.

2 The modularized structure of the framework provides an excellent opportunity for module-specific updates, which our project aims to achieve.

DeepDRR (Unberath M et al., 2018)
Cons

- 1 Only air/bone/soft tissue are segmented in current framework. And some of the segmented result(e.g. bone) isn't good enough.

- 2 Some of the simulation models need to be updated.

- 3 A lack of quantitative downstream evalutation of simulation result.

DeepDRR (Unberath M et al., 2018)

Takeaways:

- 1 The general various modules for each physical process simulation.
- 2 The general pipeline for X-ray image simulation.
- 3 Standard I/O for following updates.

Paper #3 –

Article | Open Access | Published: 04 August 2021

Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance

Manuel Schultheiss, Philipp Schmette, Jannis Bodden, Juliane Aichele, Christina Müller-Leisse, Felix G. Gassert, Florian T. Gassert, Joshua F. Gawlitza, Felix C. Hofmann, Daniel Sasse, Claudio E. von Schacky, Sebastian Ziegelmayer, Fabio De Marco, Bernhard Renger, Marcus R. Makowski, Franz Pfeiffer & Daniela Pfeiffer

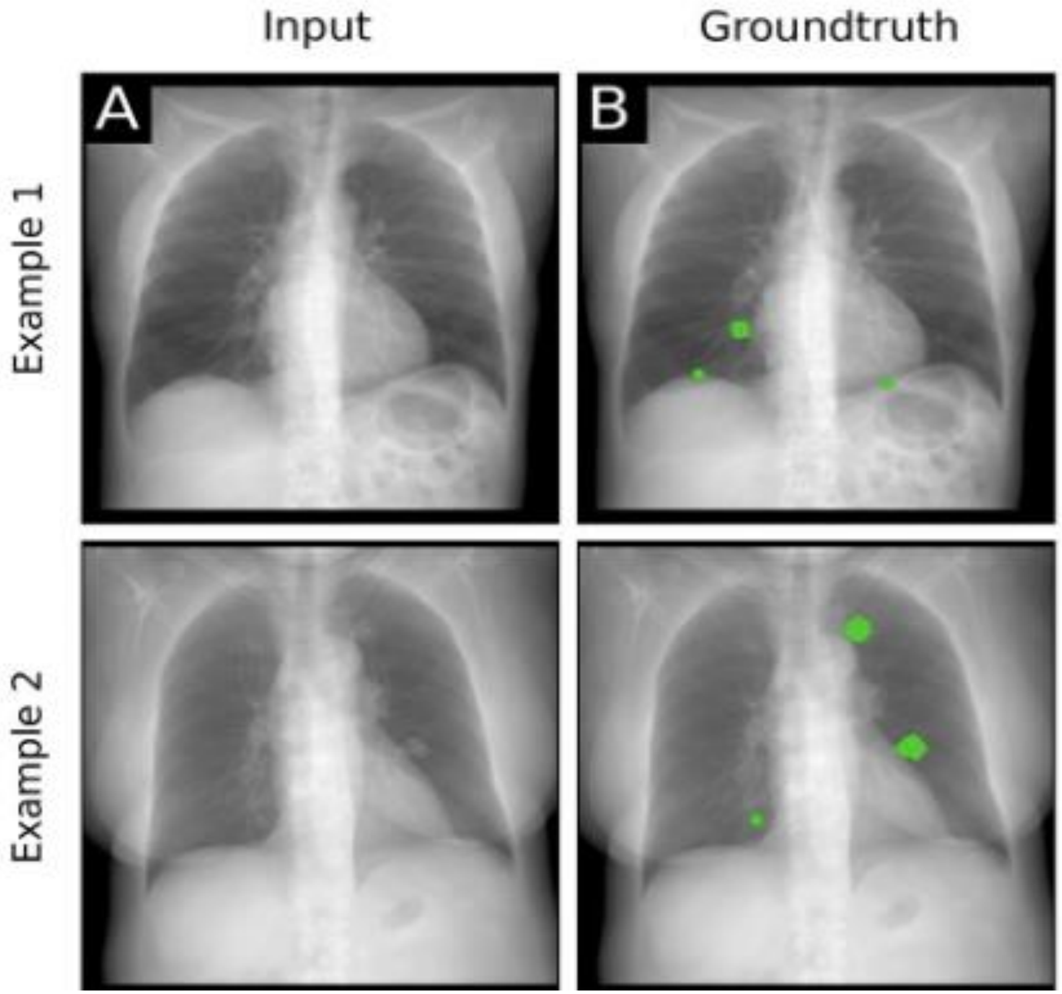
Scientific Reports **11**, Article number: 15857 (2021) | Cite this article

2656 Accesses | **2** Citations | Metrics

Reason for selection

- Our project requires a carefully constructed method of evaluation
- We model the methods described in this paper to compare the performance of A.I. models trained on either real data, current DeepDRRs, or our improved DeepDRRs

Paper #3 – Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance



Quick Abstract

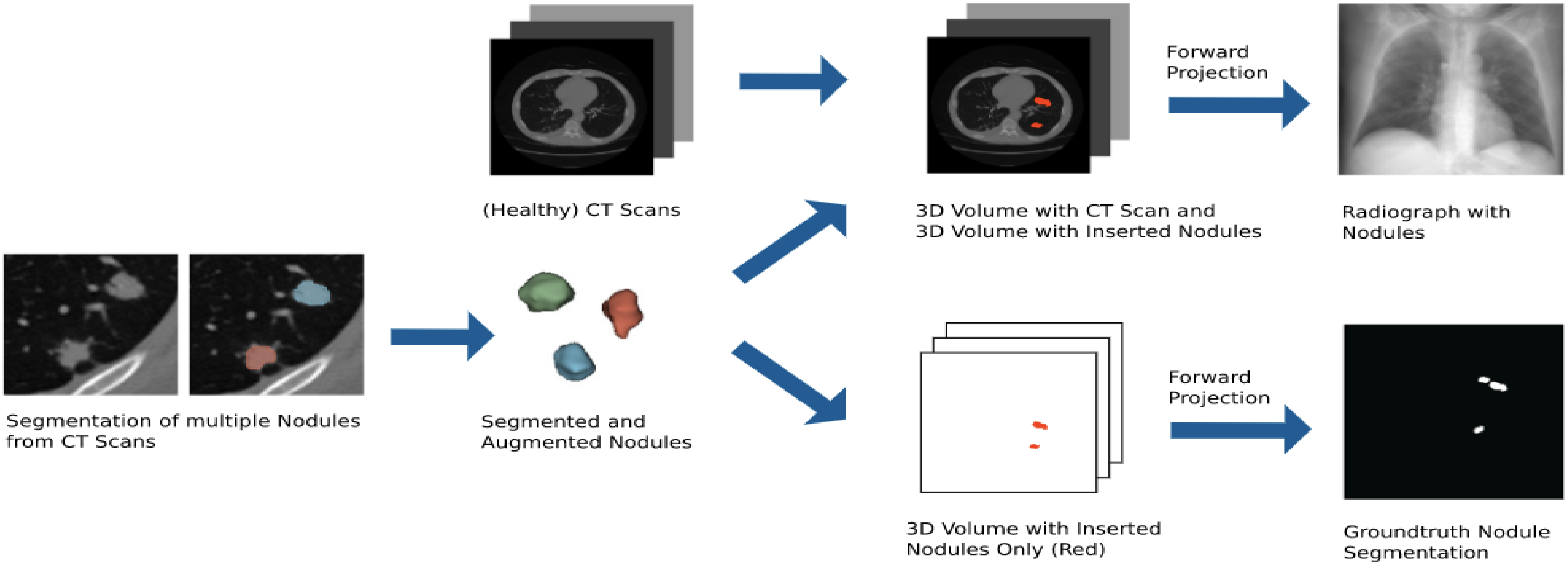
- Proposes method to randomly insert nodules into CT scan, generate synthetic X-ray images to train CNN for nodule detection task
- Compares CNN performance to that of trained radiologists

Key question:

- Is this a valid method to evaluate the quality of synthetic data?
- Is this method generalizable to other downstream tasks?

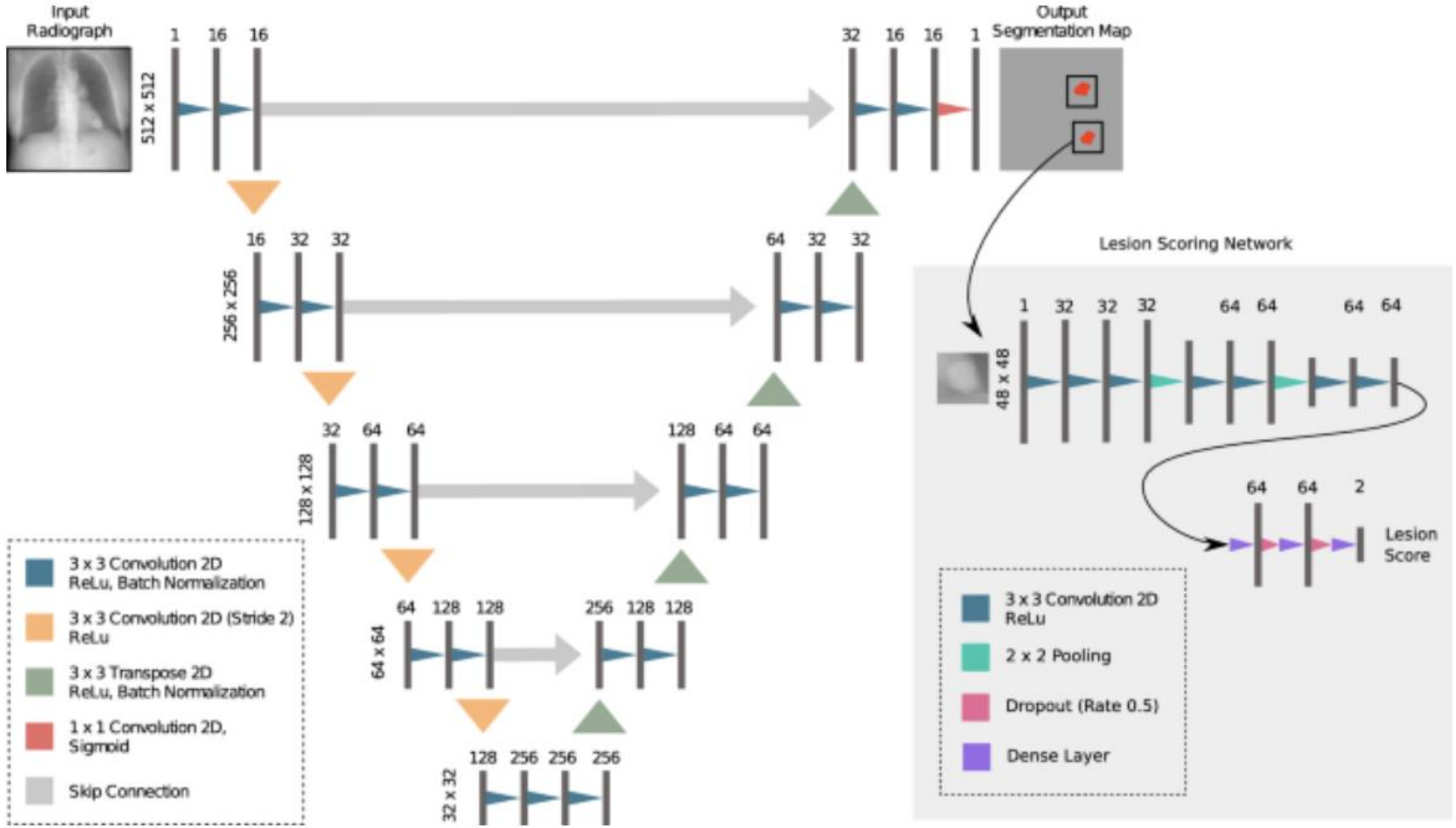
Source: Schultheiss, M., Schmette, P., Bodden, J., Aichele, J., Müller-Leisse, C., Gassert, F. G., ... & Pfeiffer, D. (2021). Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance. *Scientific Reports*, *11*(1), 1-10.

Pulmonary Nodule Detection: Ground-truth Generation



Source: Schultheiss, M., Schmette, P., Bodden, J., Aichele, J., Müller-Leisse, C., Gassert, F. G., ... & Pfeiffer, D. (2021). Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance. *Scientific Reports*, *11*(1), 1-10.

Pulmonary Nodule Detection: U-Net + Lesion Scoring Network



Source: Schultheiss, M., Schmette, P., Bodden, J., Aichele, J., Müller-Leisse, C., Gassert, F. G., ... & Pfeiffer, D. (2021). Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance. *Scientific Reports*, *11*(1), 1-10.

Paper #3 – Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance

Pros

1. This paper provides a generalizable process for evaluating the clinical applicability of simulated X-ray data. This nodule detection task may be applied to additional organs segmented in our updated DeepDRR framework, such as the liver.

ARCADE LABORATORY FOR Computational Sensing + Robotics

Paper #3 – Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance

Cons

1. As discussed with our mentors, 'sim-to-real' effects which currently limit the clinical applicability of simulated data may 'overshadow' any improvements that our 3D segmentation module may provide. This would manifest as little-to-no improvement in the performance of the updated DeepDRR-trained U-net + lesion scoring network as compared to that of current DeepDRR-trained network.

2. An additional pre-processing step is required to insert nodules into healthy CT scans for ground truth generation. This requires more compute than simpler tasks such as landmark detection in bones which has been used to evaluate DeepDRR previously [2].

ARCADE

LABORATORY FOR Computational Sensing + Robotics

Paper #3 – Lung nodule detection in chest X-rays using synthetic ground-truth data comparing CNN-based diagnosis to human performance

Key Takeaways:

1. We will start with a lung nodule segmentation task as inspired by [3] to evaluate the improved clinical applicability of our updated DeepDRRs.

2. We will generalize this nodule segmentation task to other organs segmented in our updated DeepDRR pipeline, such as the liver and kidneys.

3. In addition, we will continue to evaluate our simulated images via bone landmark detection task, as in [2].

ARCADE LABORATORY FOR Computational Sensing + Robotics