Detection and Guidance of K-Wire Placement in Pelvic Trauma Surgery

Team: Kinjal Shah and Irina Bataeva
Mentor: Dr. Ali Uneri
CIS2 Project
May 5, 2020

Motivation: K-wires in Pelvic Trauma Surgery

Motivation: Improve accuracy of K-wire detection to reduce injury while limiting radiation exposure

Objective: Use deep learning methods to:

(1) Detect K-wires in 2D radiographs of the pelvis

(2) Localize their 3D pose to provide surgical navigation

J. Goerres, et al."Deformable 3D-2D registration for guiding K-wire placement in pelvic trauma surgery", Proc. SPIE 10135, Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling, 101350A (3 March 2017);

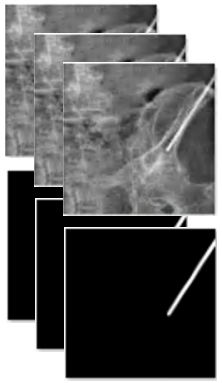
Approach

Simulated Dataset Generation

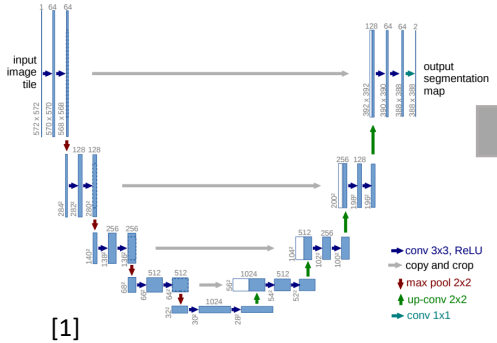
Training of Neural Network

Perform Prediction on Test Dataset

Prediction Results



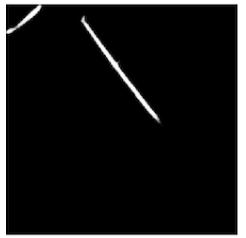
Images



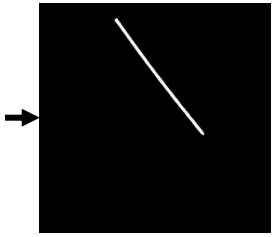
[1]

Weights

Output



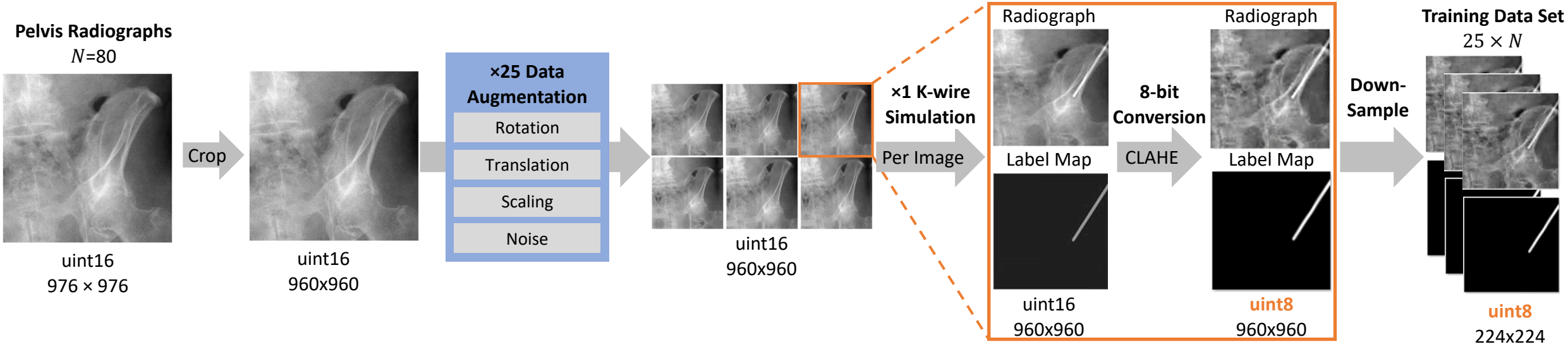
Manually annotated labels



Analysis of workflow performance

[1] Ronneberger, Olaf, et al. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” *Lecture Notes in Computer Science Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241., doi:10.1007/978-3-319-24574-4_28.

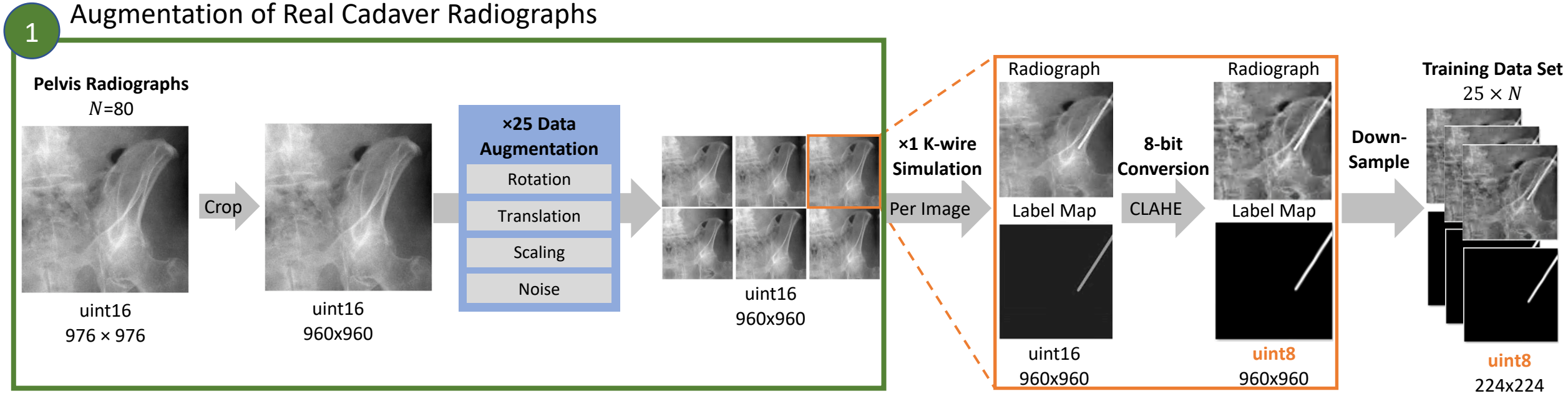
Dataset Generation: Transfer Learning



Access to quality, annotated training datasets remains a pain point for deep learning based medical image analysis

We propose a transfer learning approach based on fully simulated training data

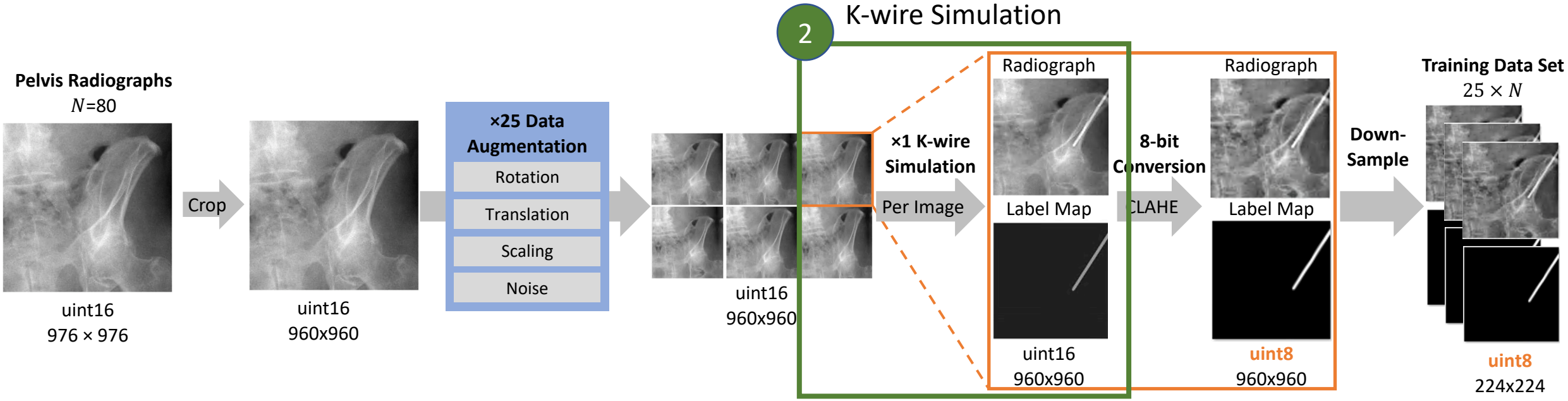
Dataset Generation: Transfer Learning



Access to quality, annotated training datasets remains a pain point for deep learning based medical image analysis

We propose a transfer learning approach based on fully simulated training data

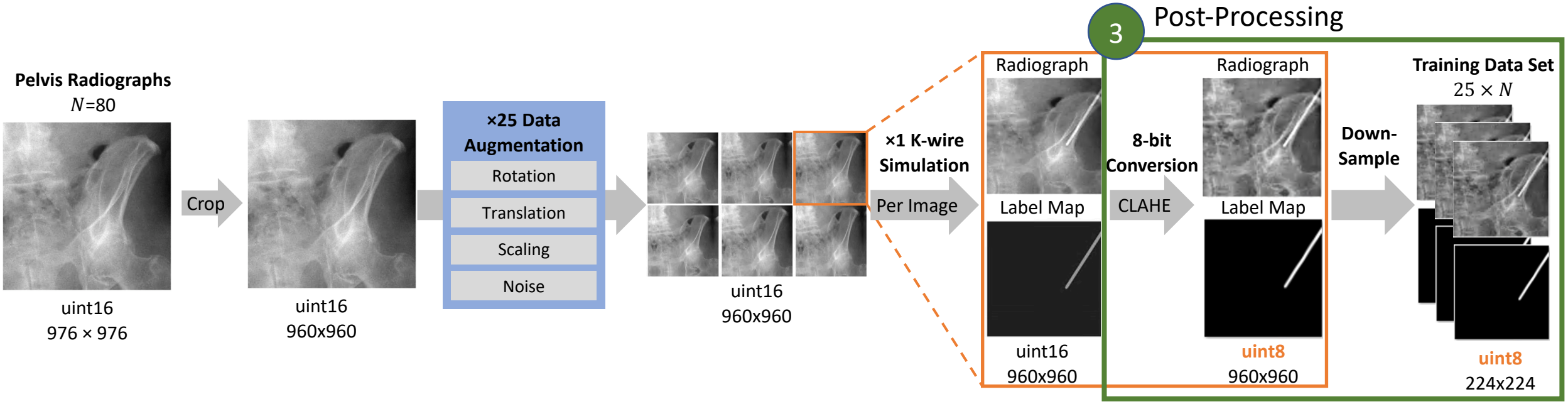
Dataset Generation: Transfer Learning



Access to quality, annotated training datasets remains a pain point for deep learning based medical image analysis

We propose a transfer learning approach based on fully simulated training data

Dataset Generation: Transfer Learning



Access to quality, annotated training datasets remains a pain point for deep learning based medical image analysis

We propose a transfer learning approach based on fully simulated training data

Training Dataset Generation: Examples

AP-like views

Single K-wire

Image sizes:

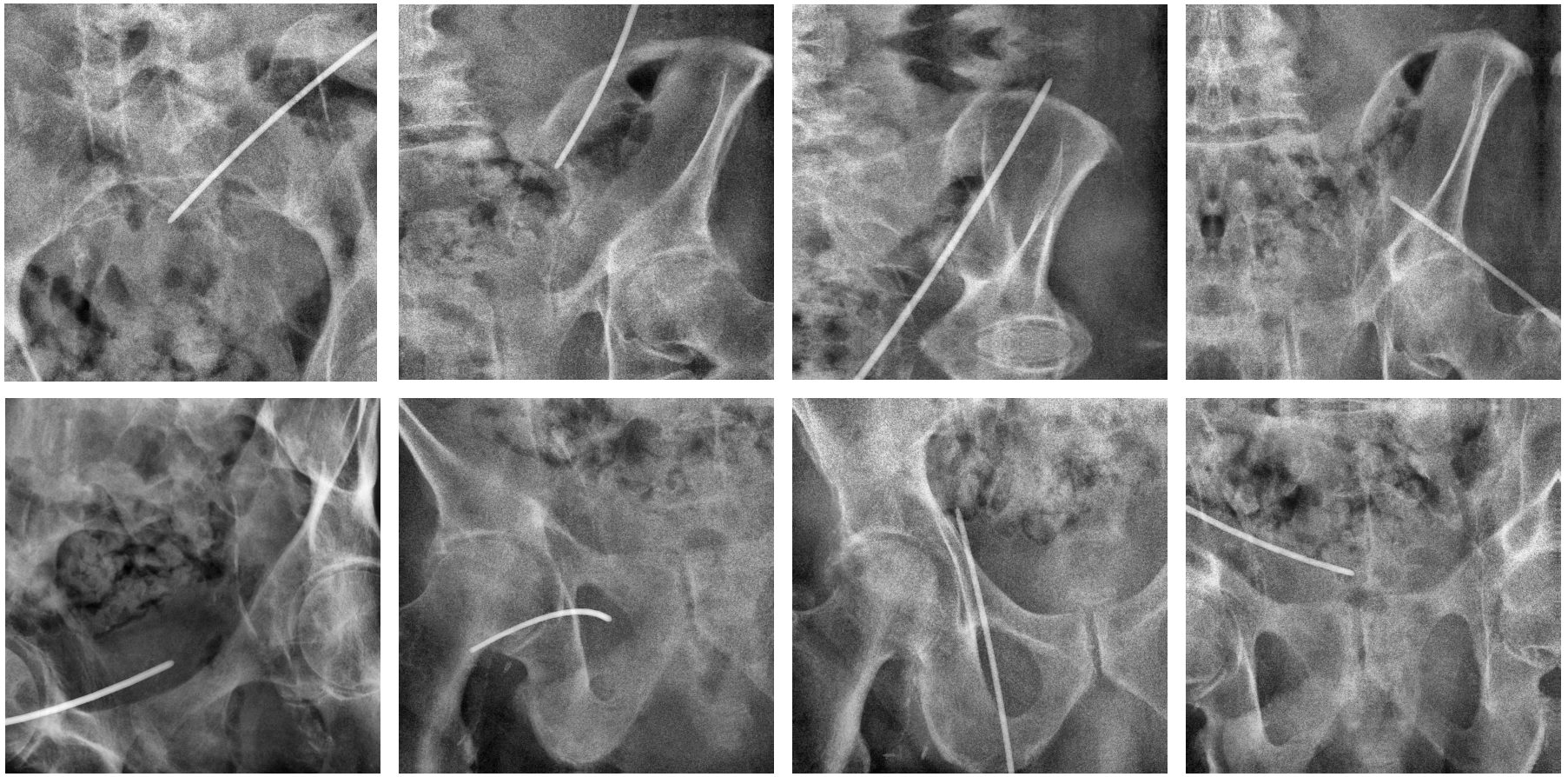
960x960
480x480
224x224

Dataset sizes:

250
500
1000
2000
4000

K-wire projection:

Deformed
Rigid



Test Dataset

IRB Approved Dataset

Acquired using Cios Spin (Siemens, Erlangen Germany)

AP-like views, single K-wire

Raw Test Dataset:

Image Size: 976x976

7 Single K-wire Images

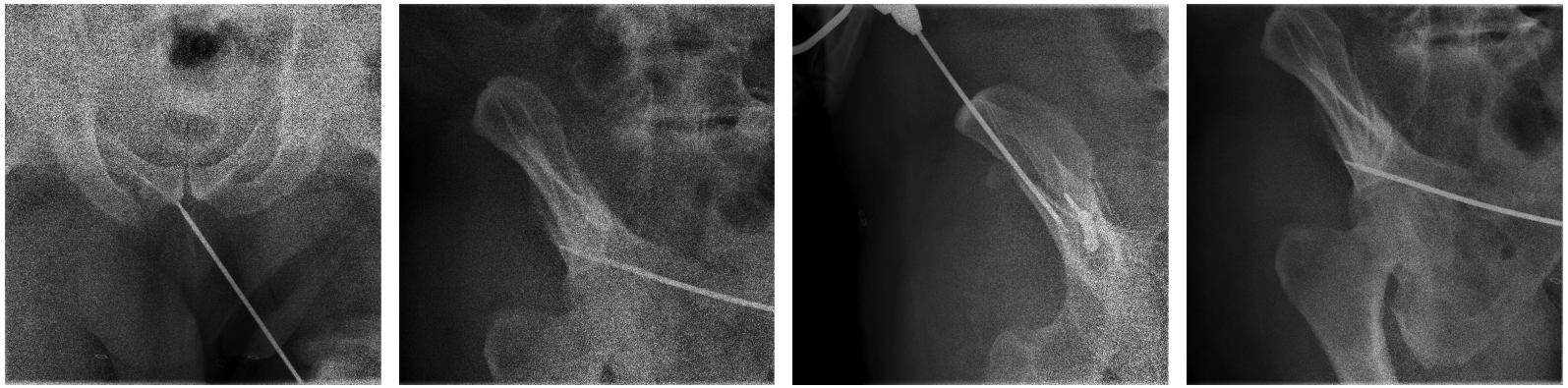
13 Multiple K-wire images

Test Dataset:

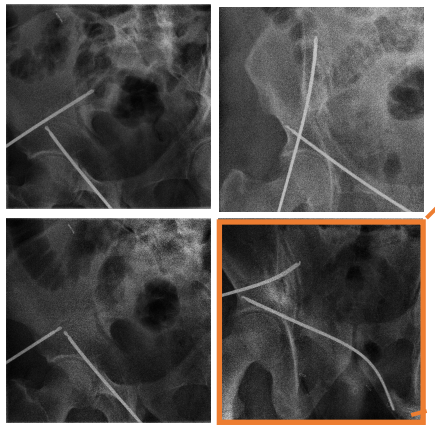
Image Size: 960x960

33 Single K-wire Images

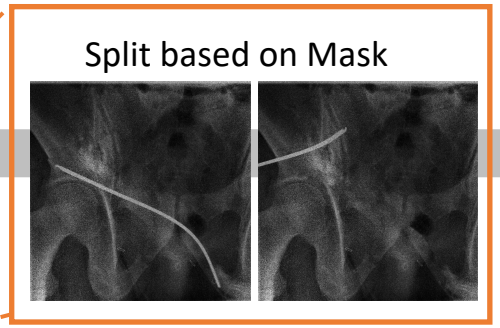
7 Single



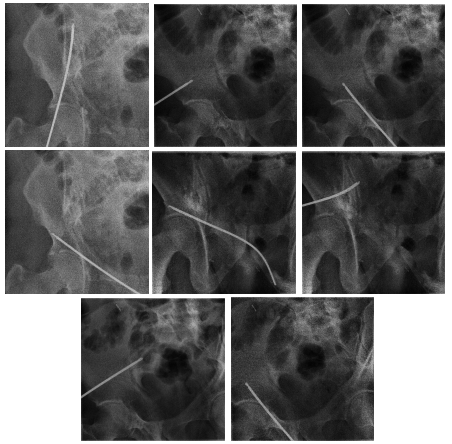
13 Multiple



Split based on Mask



26 Single



Figures of Merit: Standard CNN Prediction Metrics

Area under the Curve

$$\text{AUC} = \text{Area under the TP Rate vs. FP Rate Curve}$$

Precision

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}}$$

Recall

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}}$$

F1 Score (Sørensen–Dice coefficient)

$$f1\ \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

		Ground Truth	
		P	N
Prediction	P	TP	FP
	N	FN	TN

Figures of Merit: Standard CNN Prediction Metrics

Area under the Curve

$$\text{AUC} = \text{Area under the TP Rate vs. FP Rate Curve}$$

Precision

$$\text{precision} = \frac{\text{TP}}{\text{TP+FP}}$$

Recall

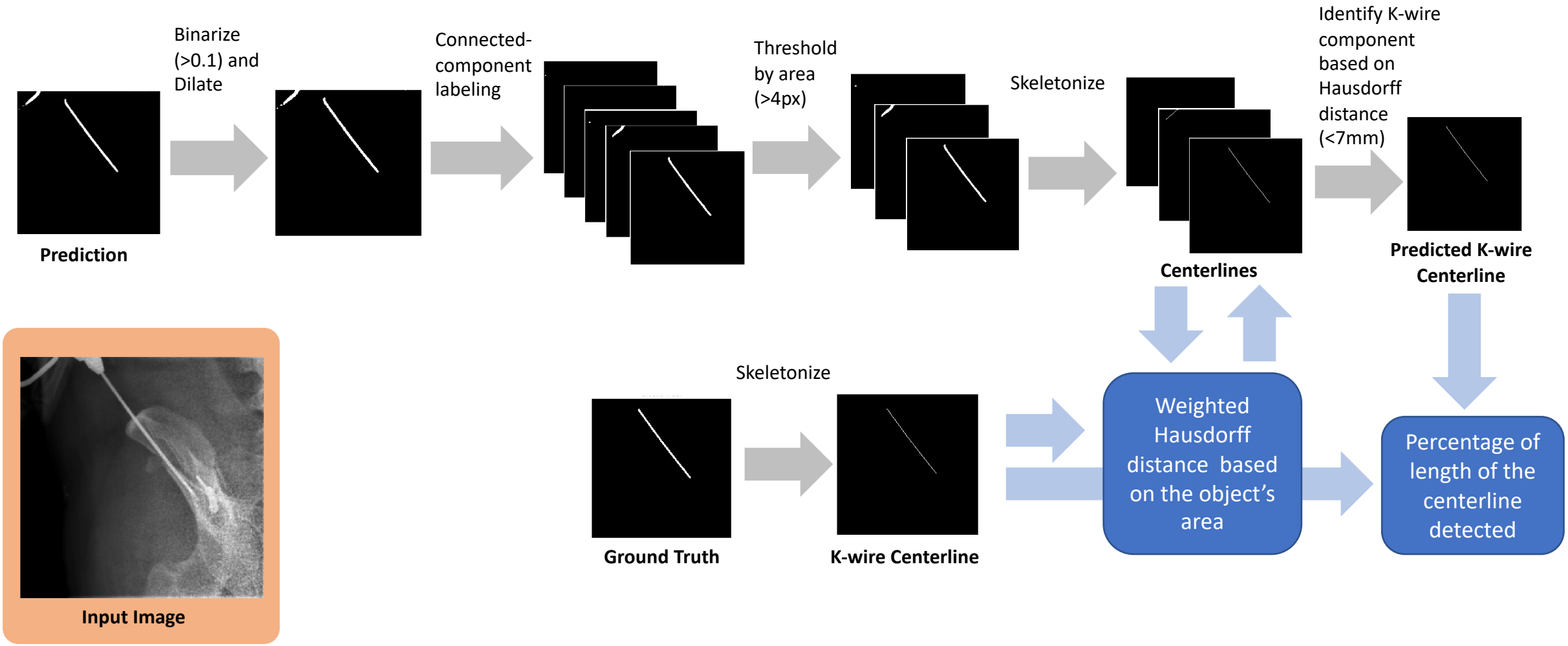
$$\mathbf{recall} = \frac{\mathbf{TP}}{\mathbf{TP+FN}}$$

F1 Score (Sørensen–Dice coefficient)

$$f1\ \text{score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

		Ground Truth	
		P	N
Prediction	P	TP	FP
	N	FN	TN

Figures of Merit: K-Wire Detection Task Specific



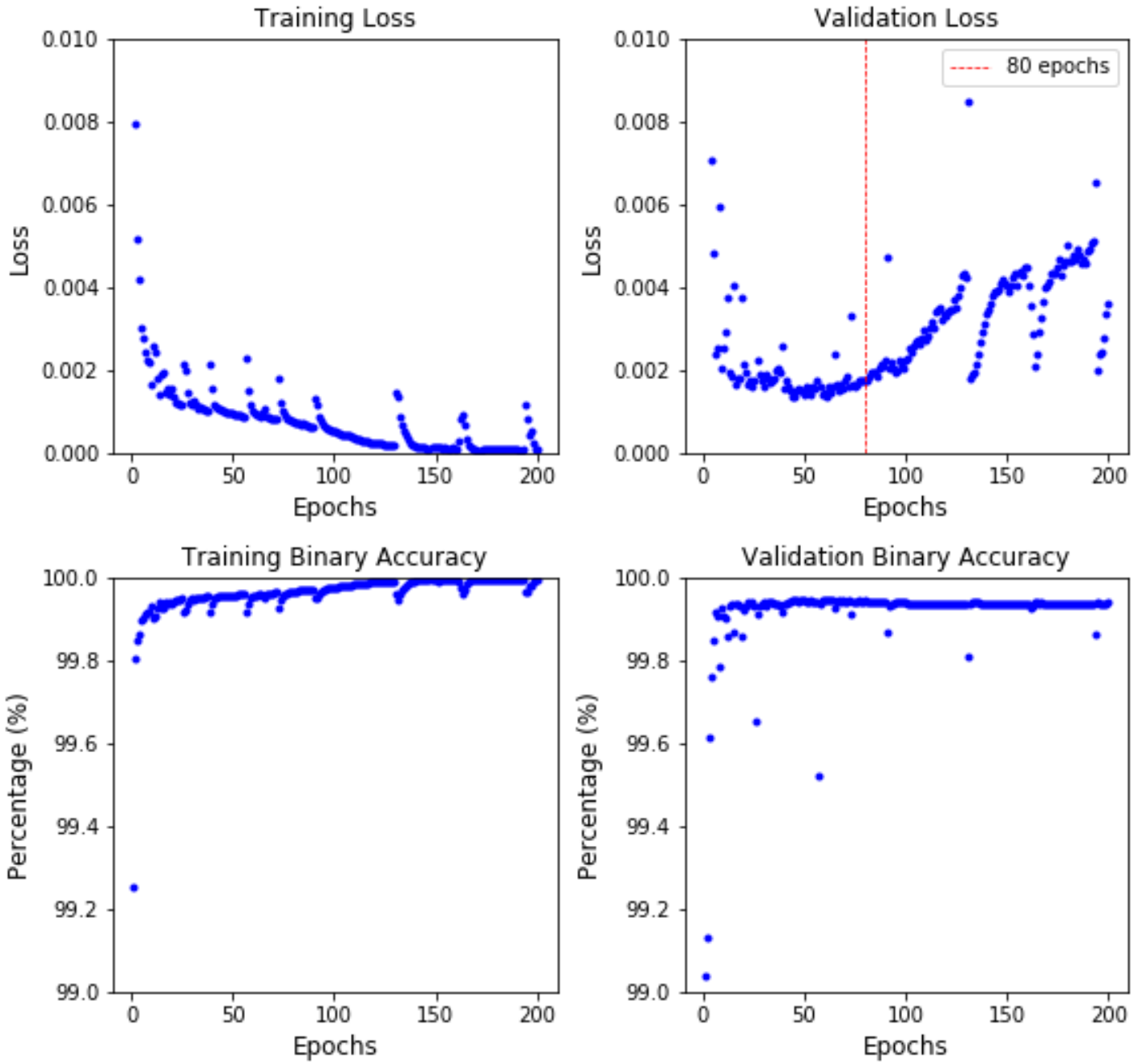
Training Time

Binary Cross-Entropy Loss Function

$$L(y, \hat{y}) = -\frac{1}{N}(y * \log(\hat{y}_i) + (1 - y) * \log(1 - \hat{y}_i))$$

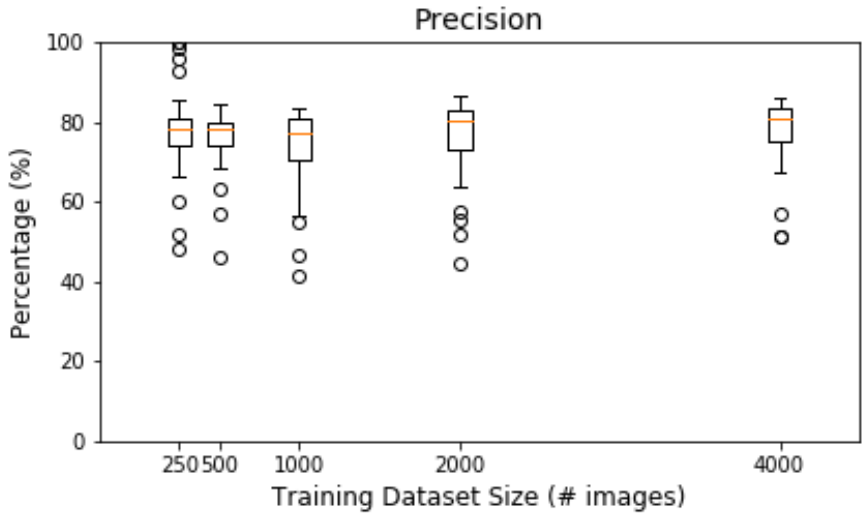
$$\text{Binary Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

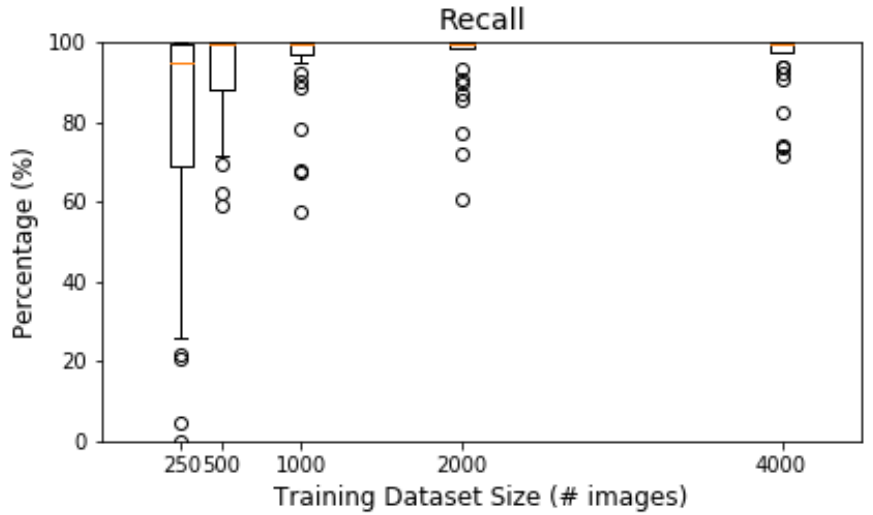
Training Dataset Size: 2000
Batch Size: 10
Epoch: 80
Image Size: 224x224

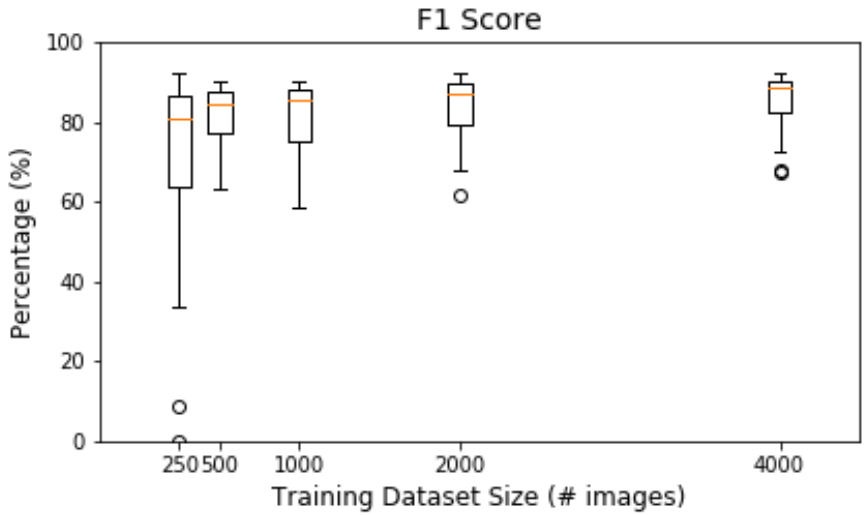


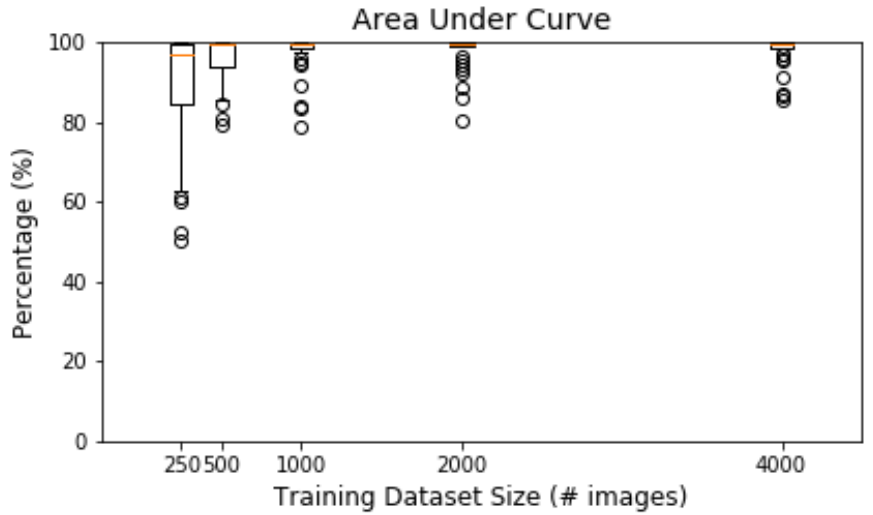
Training Dataset Size

General improvement as training dataset size increases







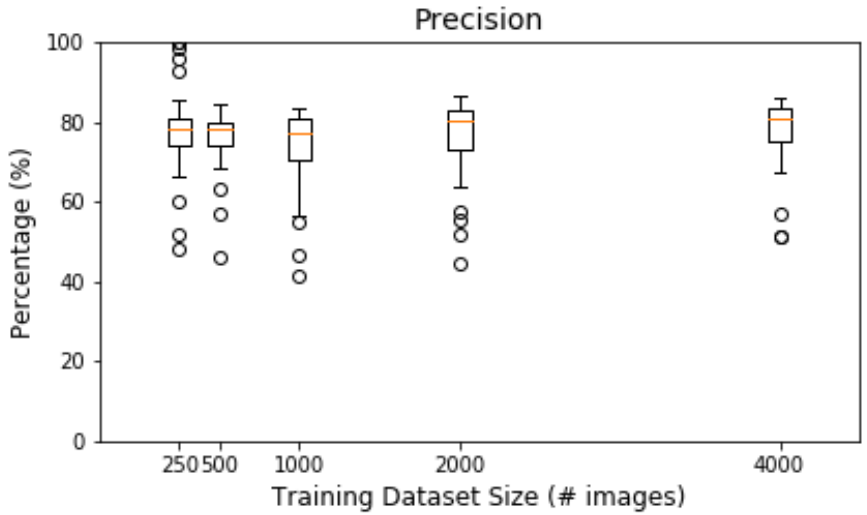


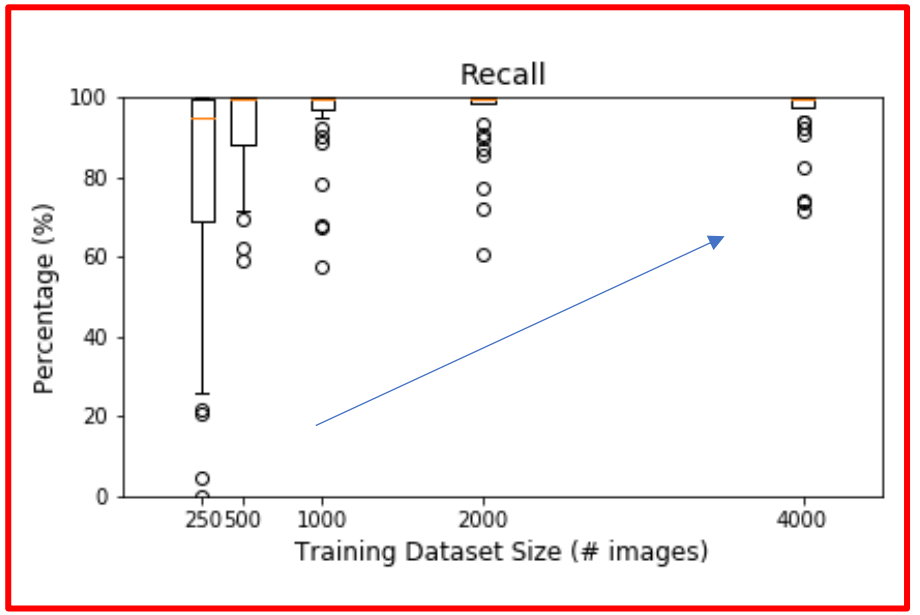
Training Dataset Size: 250–4000
Batch Size: 10
Epoch: 80
Image Size: 224x224

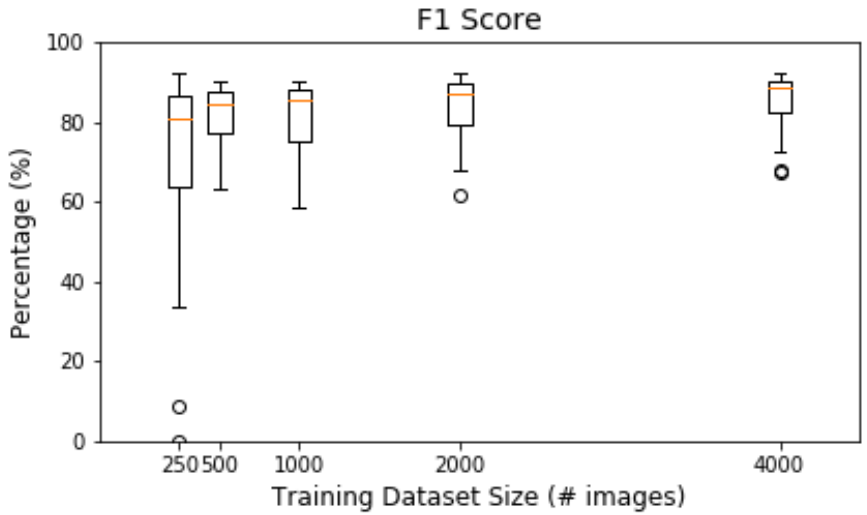
Training Dataset Size

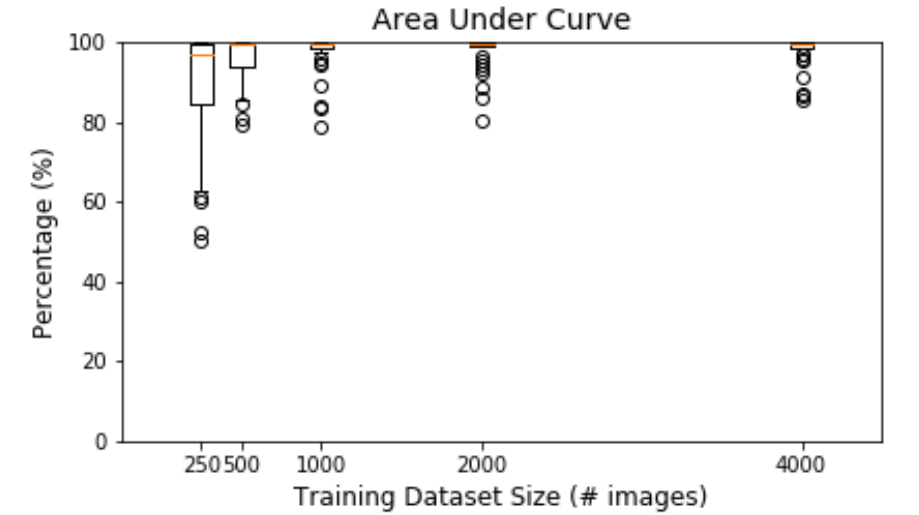
Recall improvement is of particular interest

Hope to see false negatives decrease



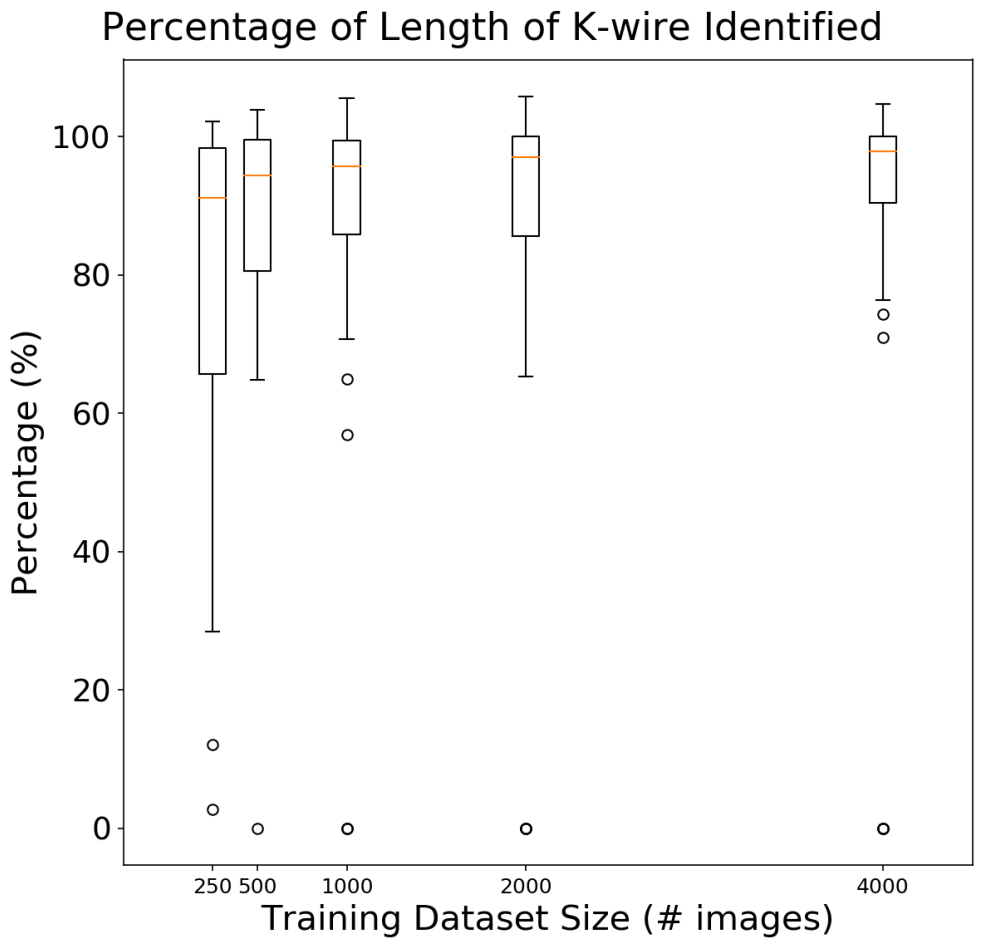




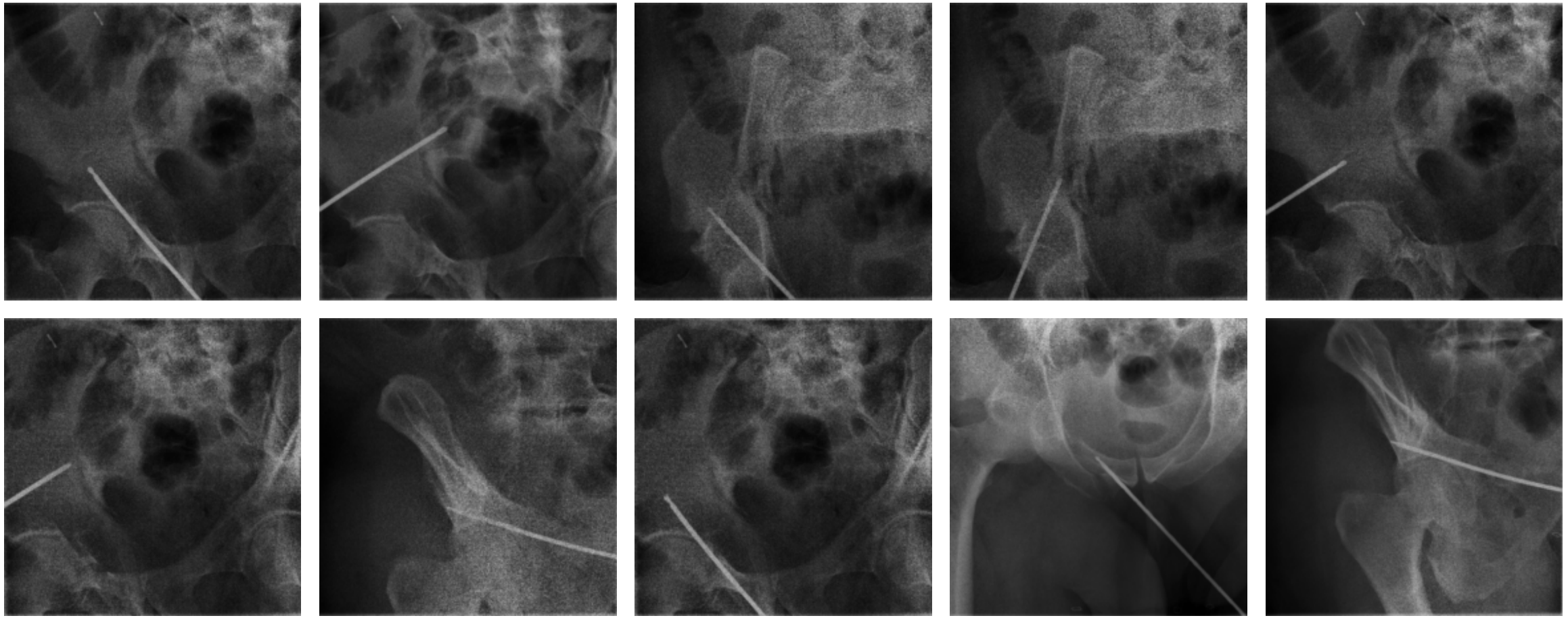


Training Dataset Size: 250–4000
Batch Size: 10
Epoch: 80
Image Size: 224x224

Training Dataset Size: Centerline Metrics



Prediction Examples



Prediction Examples

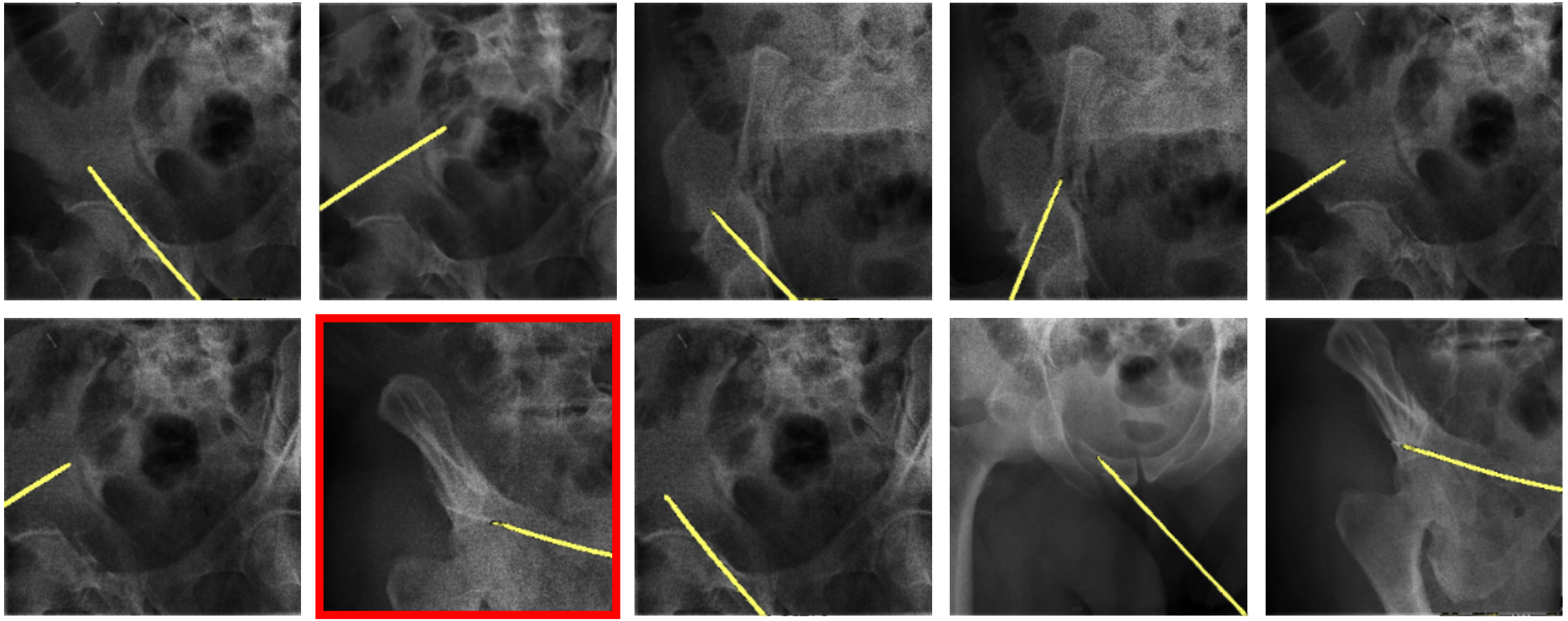
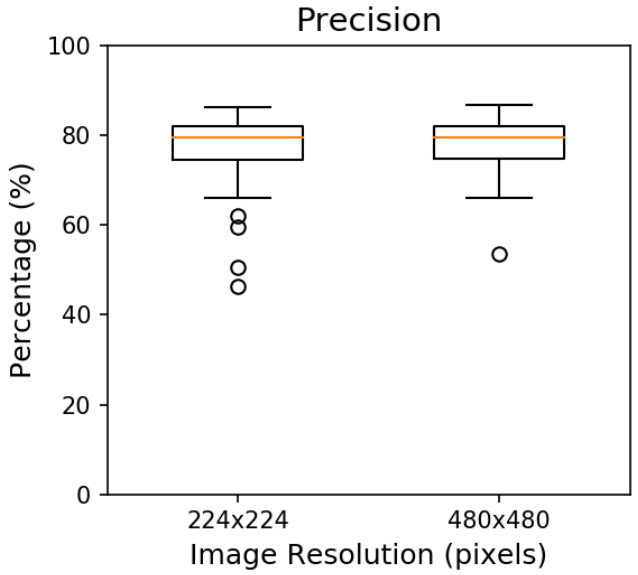
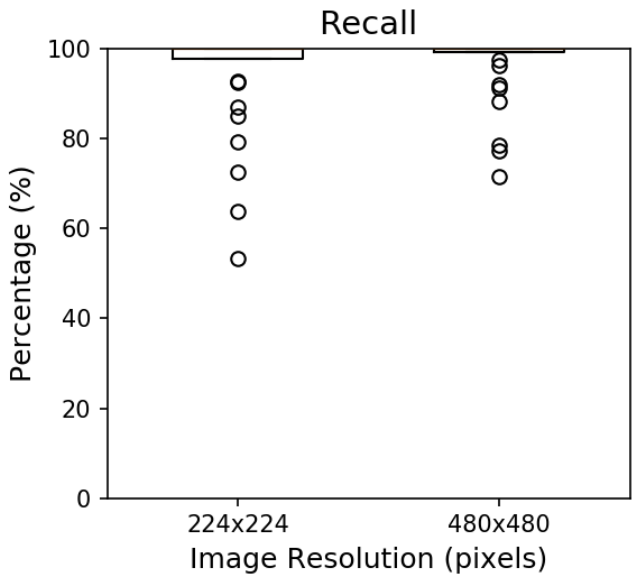


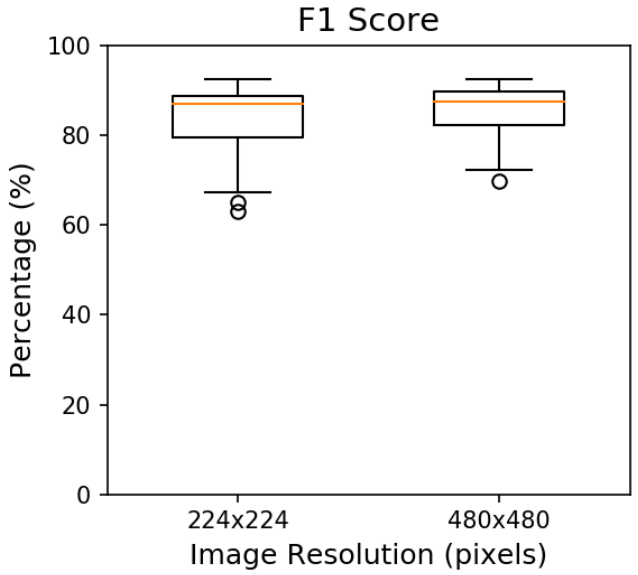
Image Resolution

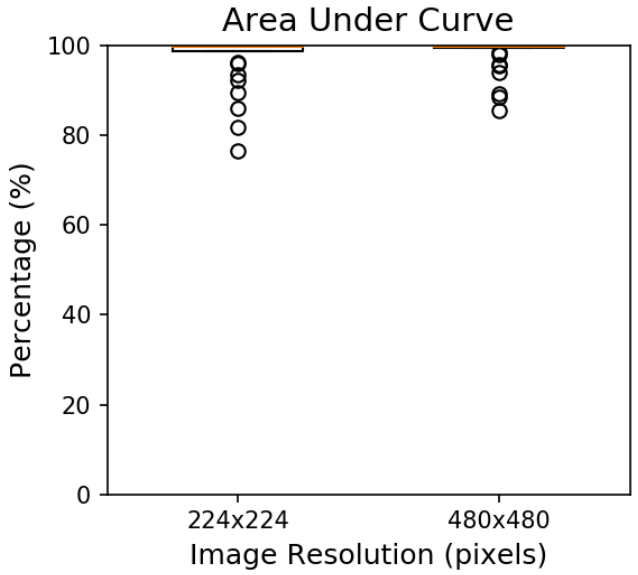
Marginal improvement in performance

*need larger test set and statistical analysis to assess





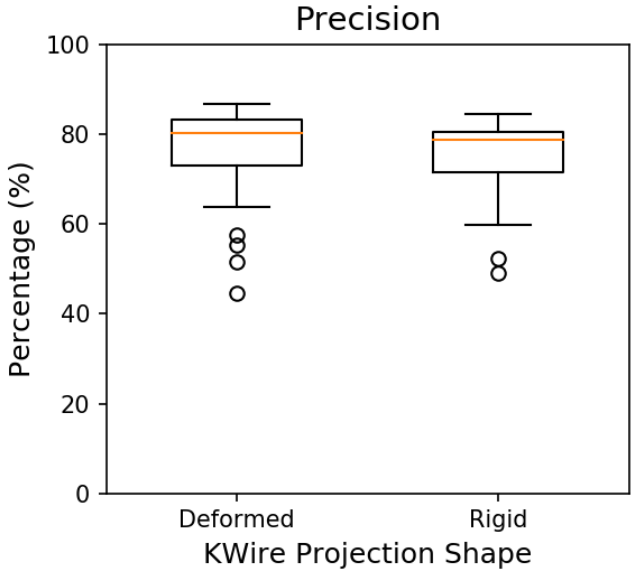


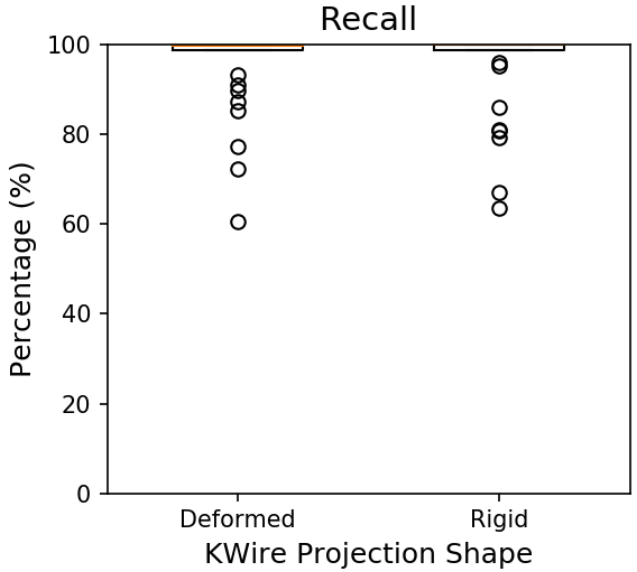


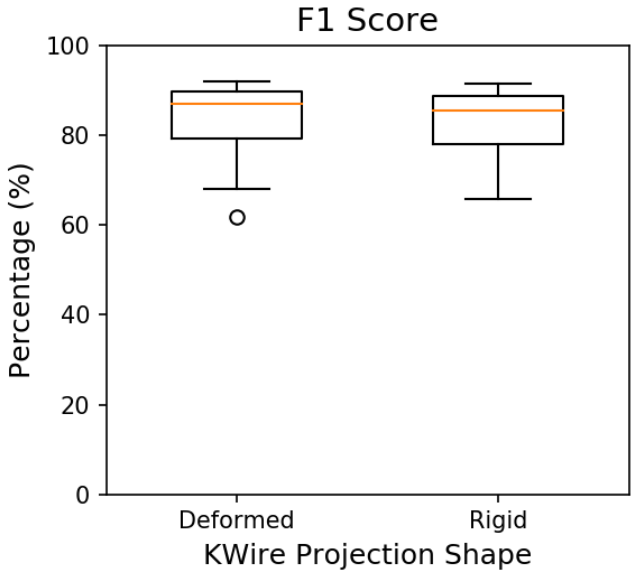
Training Dataset Size: 2000
Batch Size: 4
Epoch: 80
Image Size: 224 vs. 480 px

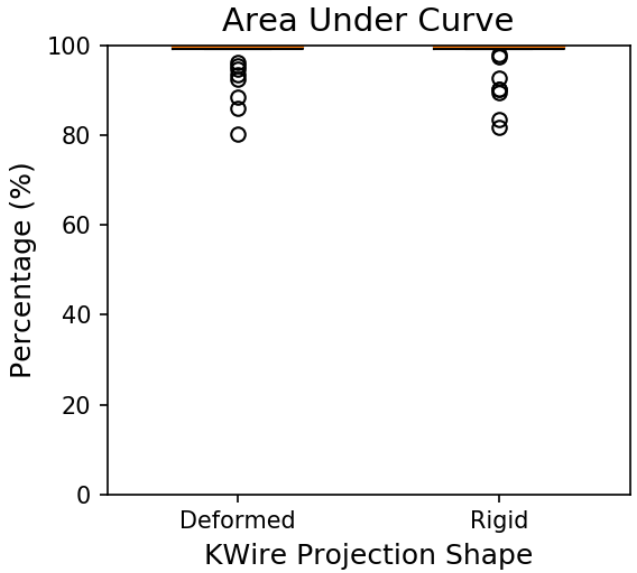
Training Dataset Simulated K-Wire Shape

Performance is essentially equivalent





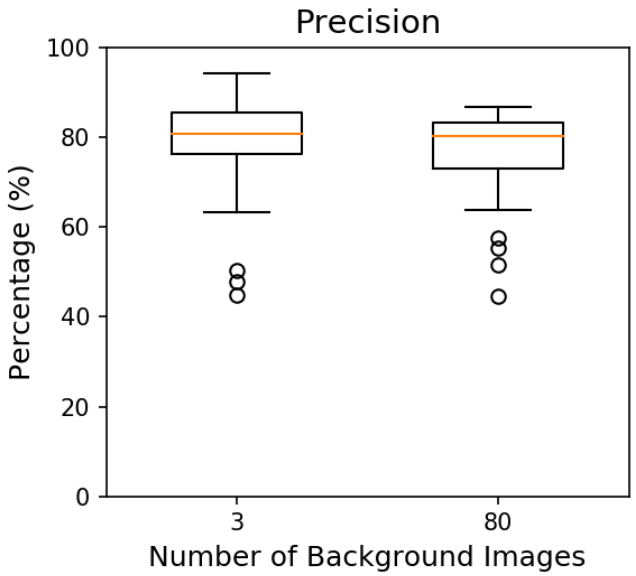


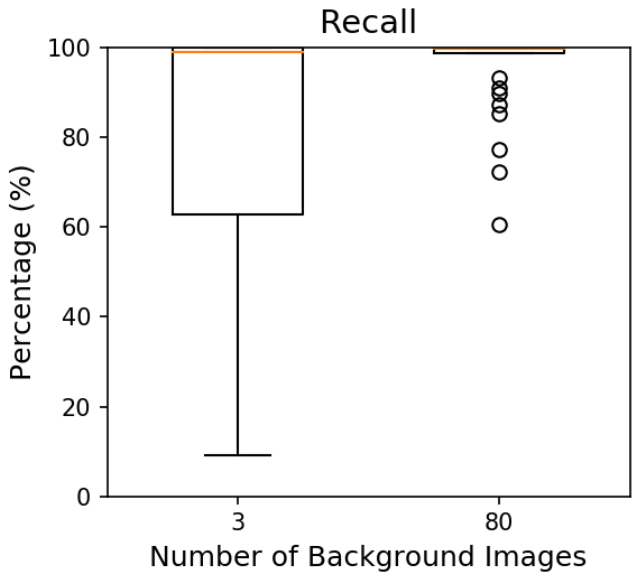


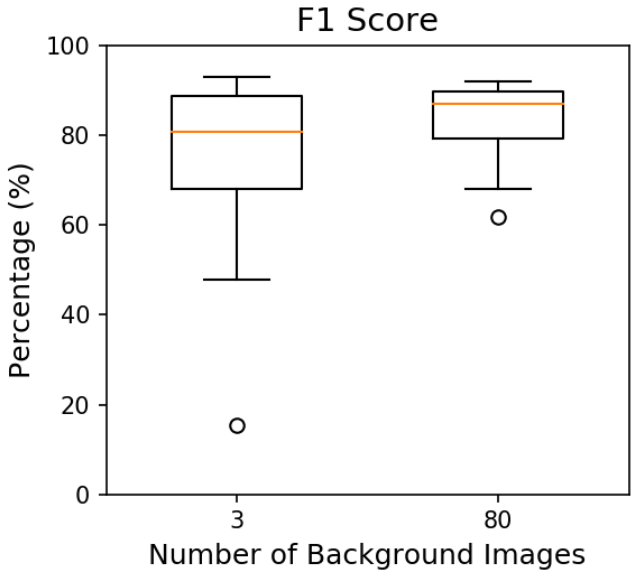
Training Dataset Size: 2000
Batch Size: 10
Epoch: 80
Image Size: 224x224

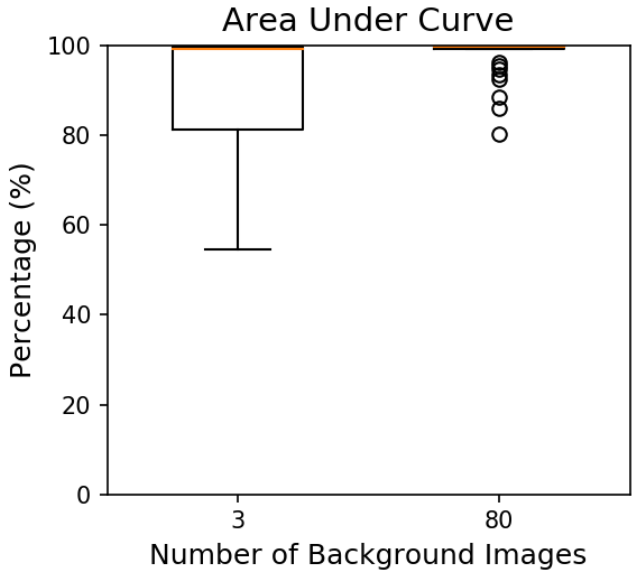
Training Dataset: Diversity of Background Image Selection

Increased diversity in background images prior to augmentation yields better results









Training Dataset Size: 2000
Batch Size: 10
Epoch: 80
Image Size: 224x224

CNN Architectures

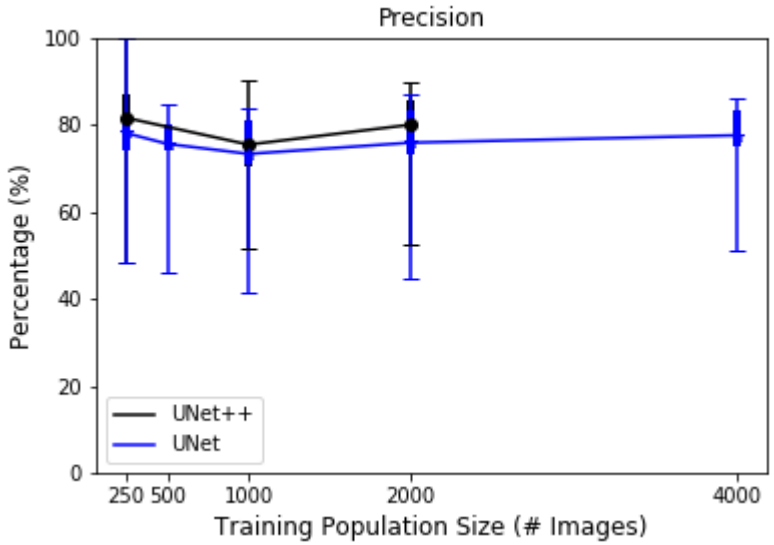
Displayed comparison

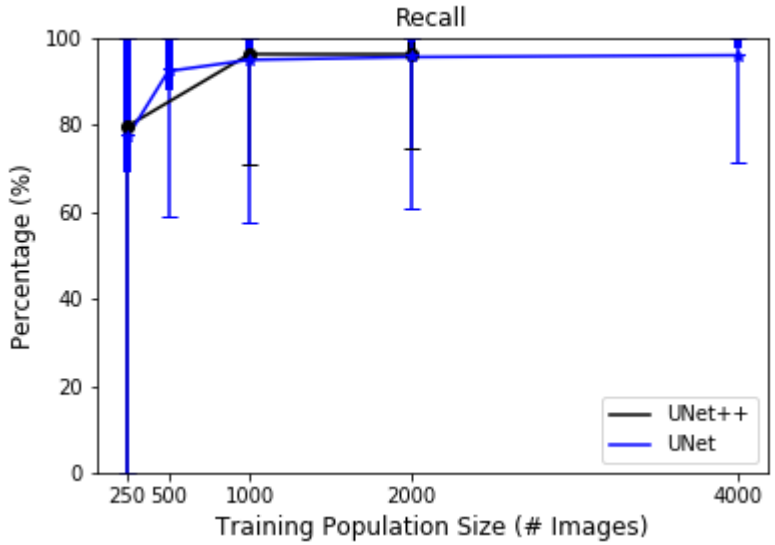
UNet to UNet++

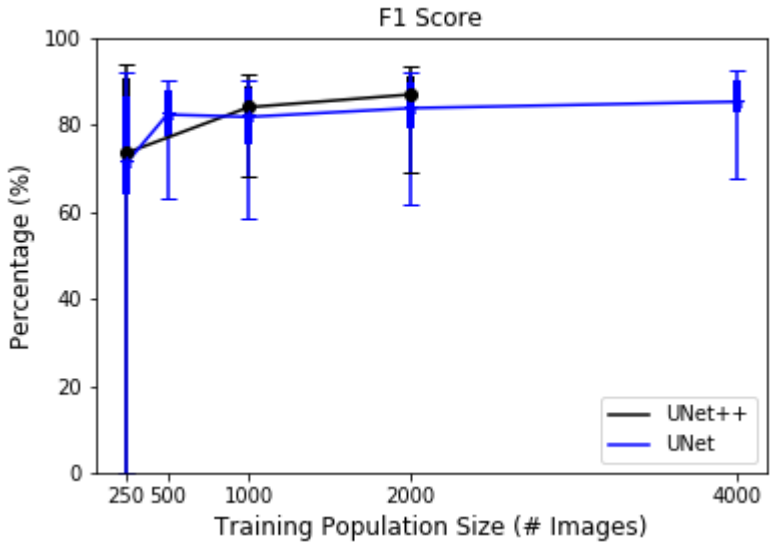
UNet++: more complex network

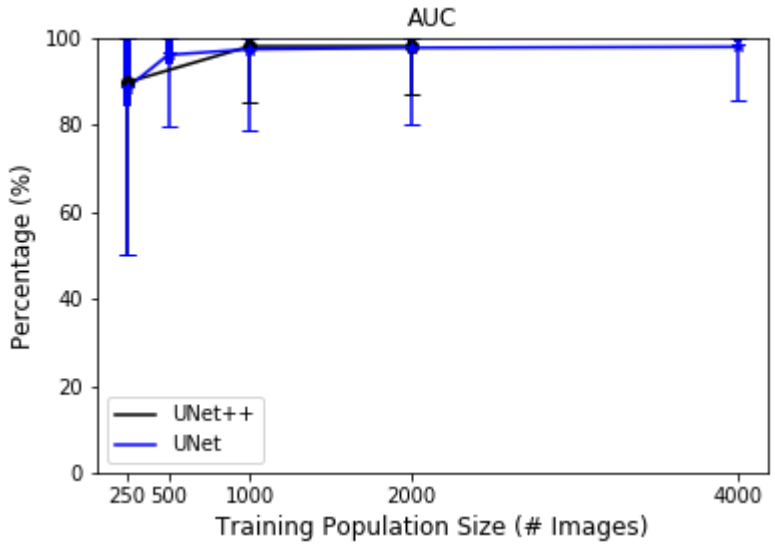
~2x the inference time / image

Similar performance









Training Dataset Size: 250–4000
Batch Size: 4
Epoch: 80
Image Size: 224x224

Summary

1. Generated simulated dataset
2. Trained and evaluated application of UNet for K-Wire detection
3. Achieved predictions on real pelvic images from real patient experiments

Minimum Deliverable

Expected Deliverable

Responsibility Ownership

Irina Bataeva	Kinjal Shah
• Training Set Generation: K-wire simulation • Test Set Segmentation • Model Evaluation: Task Specific Metrics	• Training Set Generation: Dataset augmentation • Model Implementation • Model Evaluation: Standard CNN Metrics

Next Steps

Improve Transfer Learning

Refine dataset generation process

Assess alternate model architectures

Multiple K-Wires

Multi-class segmentation, object detection

3D Localization

Multi-class segmentation

Maximum Deliverable

Questions

Thank you!

Appendix: Prediction Example

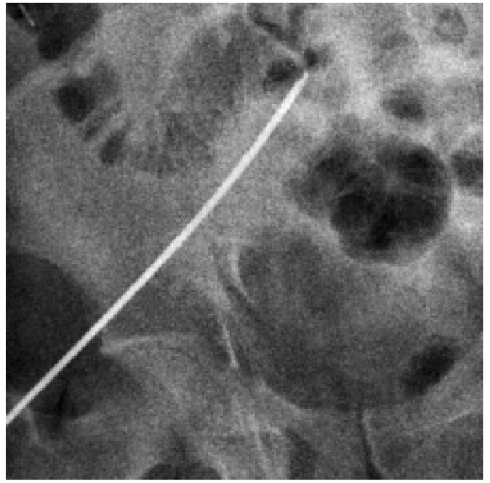
Input Image

Ground Truth

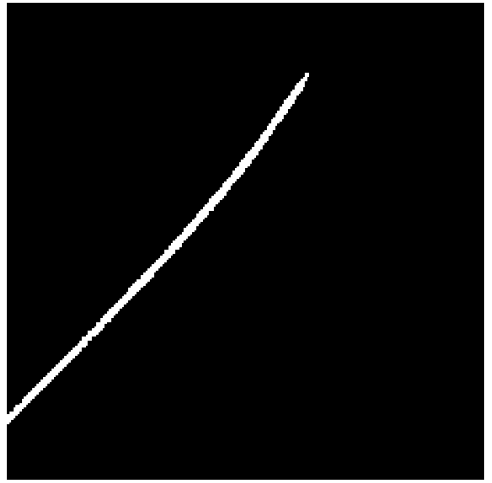
Prediction

Failure Case

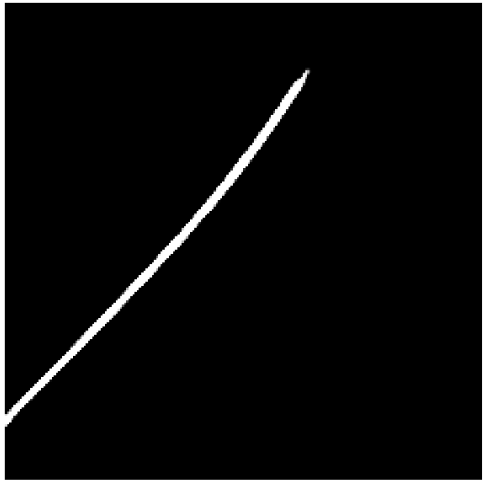
Ideal Case



Input Image

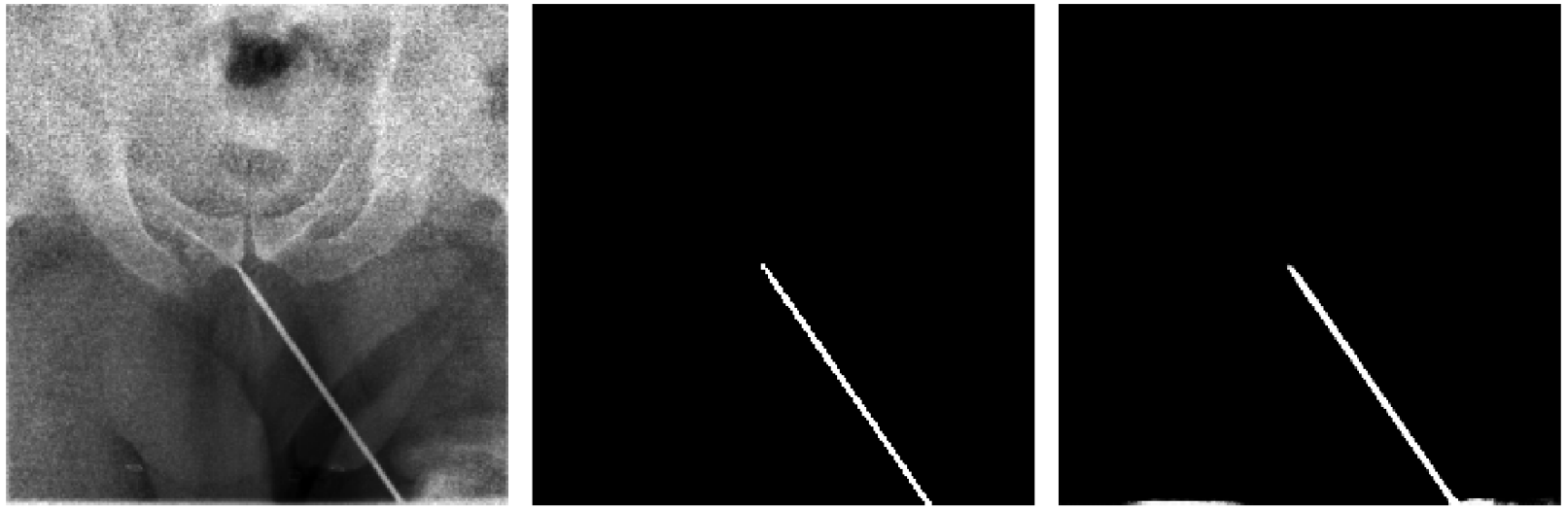


Ground Truth



Prediction

Appendix:
Prediction Example (4000 image training dataset) – Image crop issue



Appendix:
Prediction Example (4000 image training dataset) – Failure