Deep Learning for Fluoroscopic Feature Detection

Computer Integrated Surgery II

Spring, 2018

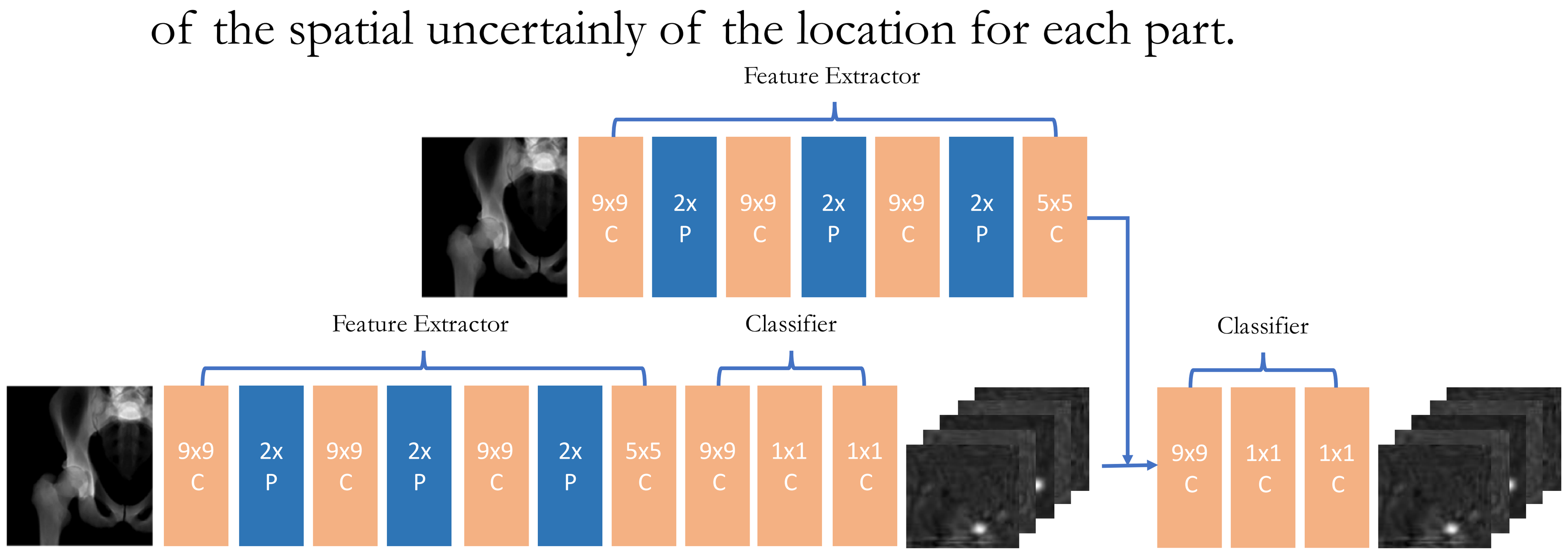
Liujiang Yan under the auspices of Robert Grupp and Professor Russell Taylor

Introduction

- Introduced convolutional pose machines for anatomical landmark detection in simulated X-Ray images;
- Introduced holistically-nested networks for contour detection in simulated X-Ray images and transferred to real X-Ray images;
- Developed random region mask as data augmentation to simulate tool in the field;
- Perform thorough and detailed experiments to evaluate the models' performance under different data.

Landmark Detection

- Convolutional pose machines consist a sequence of convolutional networks that repeatedly produce 2D belief map for each part.
- Belief maps at each stage represent the non-parametric encoding of the spatial uncertainly of the location for each part.

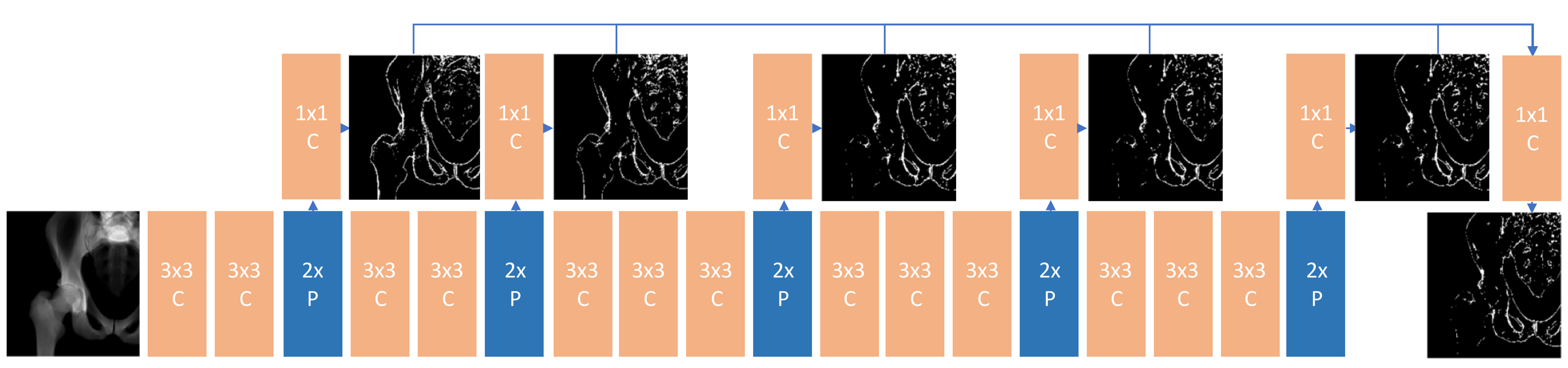


- The objective is to minimize the L2 distance defined by the belief map from each stage and the target belief map for each part.

$$F = \sum_{t=1}^{T} \sum_{p}^{P+1} \sum_{x} ||b_t^p(x) - b_{gt}^p(x)||_2^2$$

Contour Detection

- The architecture consists of a VGG convolutional net for multi-scale feature learning, and assign a side output at each stage, and a fusion layer aggregating all stages.



- The objective is to minimize class-weighted cross entropy defined at each side output and fusion output.

$$L = \sum_{t=1}^{T+1} \left(-\beta \sum_{j \in Y^+} \log(y_j = 1|X) - (1-\beta) \sum_{j \in Y^-} \log(y_j = 0|X) \right)$$

Data Augmentation

- randomly mask a region of arbitrary shape, arbitrary size, and arbitrary constant intensity to simulate tool in the field.

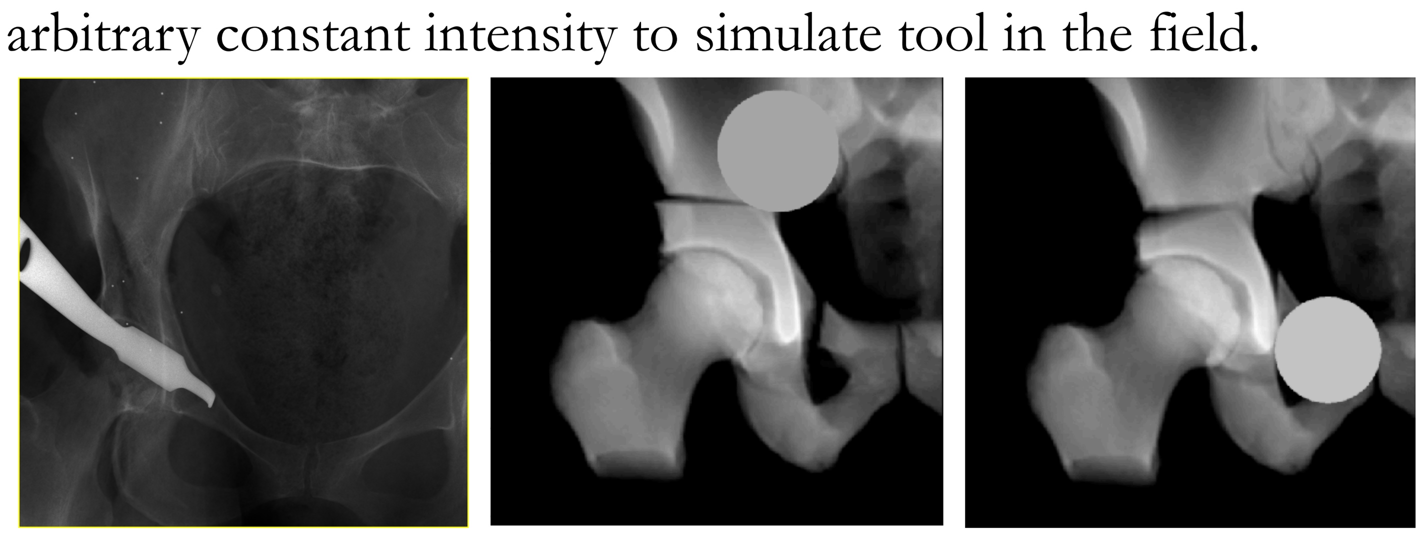


Figure 6 Random Region Mask as Data Augmentation

Future Work

- **Network Architecture**: The basic block for these networks are pure convolution and pooling layer, and could be replaced by other well-performed alternatives: residual layer, U-Net.
- **Transfer Learning**: Model trained on simulated data does not perform well on real X-Ray data. A potential approach is to fine-tune the model by real X-Ray data and then evaluate.
- **Evaluation Metric**: While the pixel distance gives an intuitive measure about the performance, the led 2D-3D transformation difference will be more straightforward.

Reference:

- Wei, Shih-En, et al. "Convolutional pose machines." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016.
- Xie, Saining, and Zhuowen Tu. "Holistically-nested edge detection." Proceedings of the IEEE international conference on computer vision. 2015.

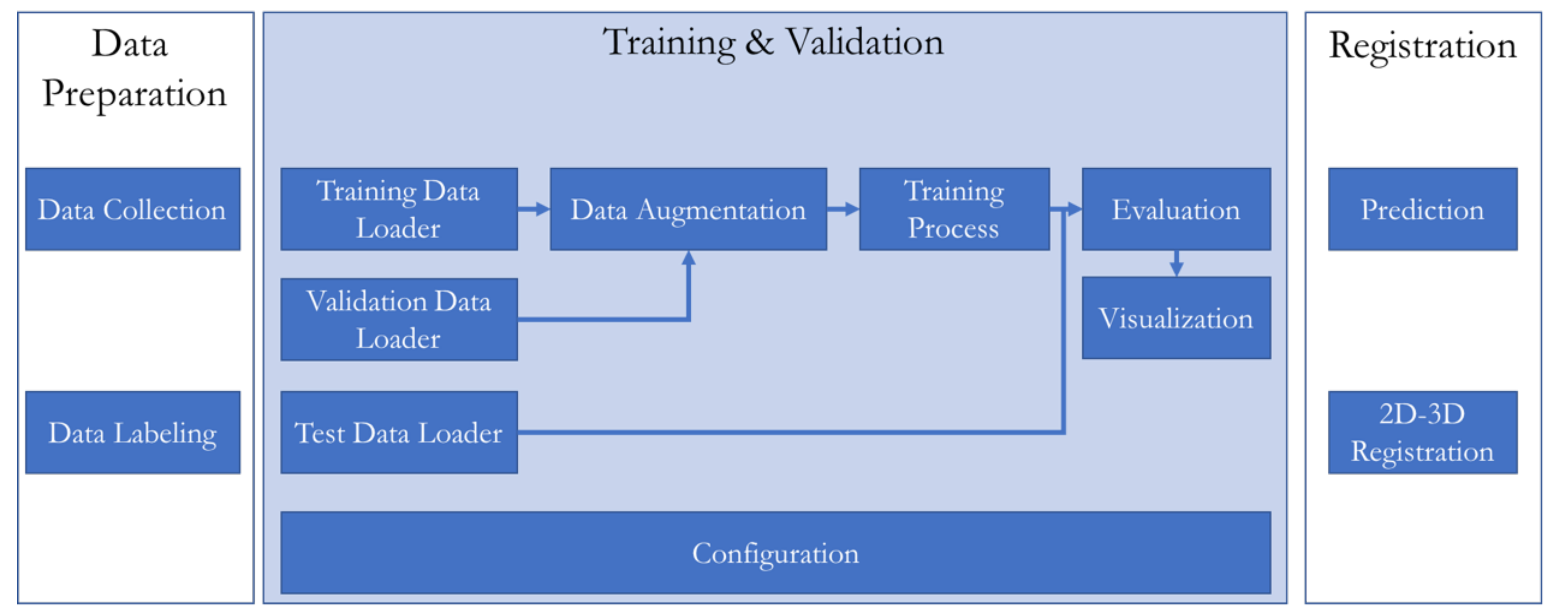


Figure 1 System Pipeline

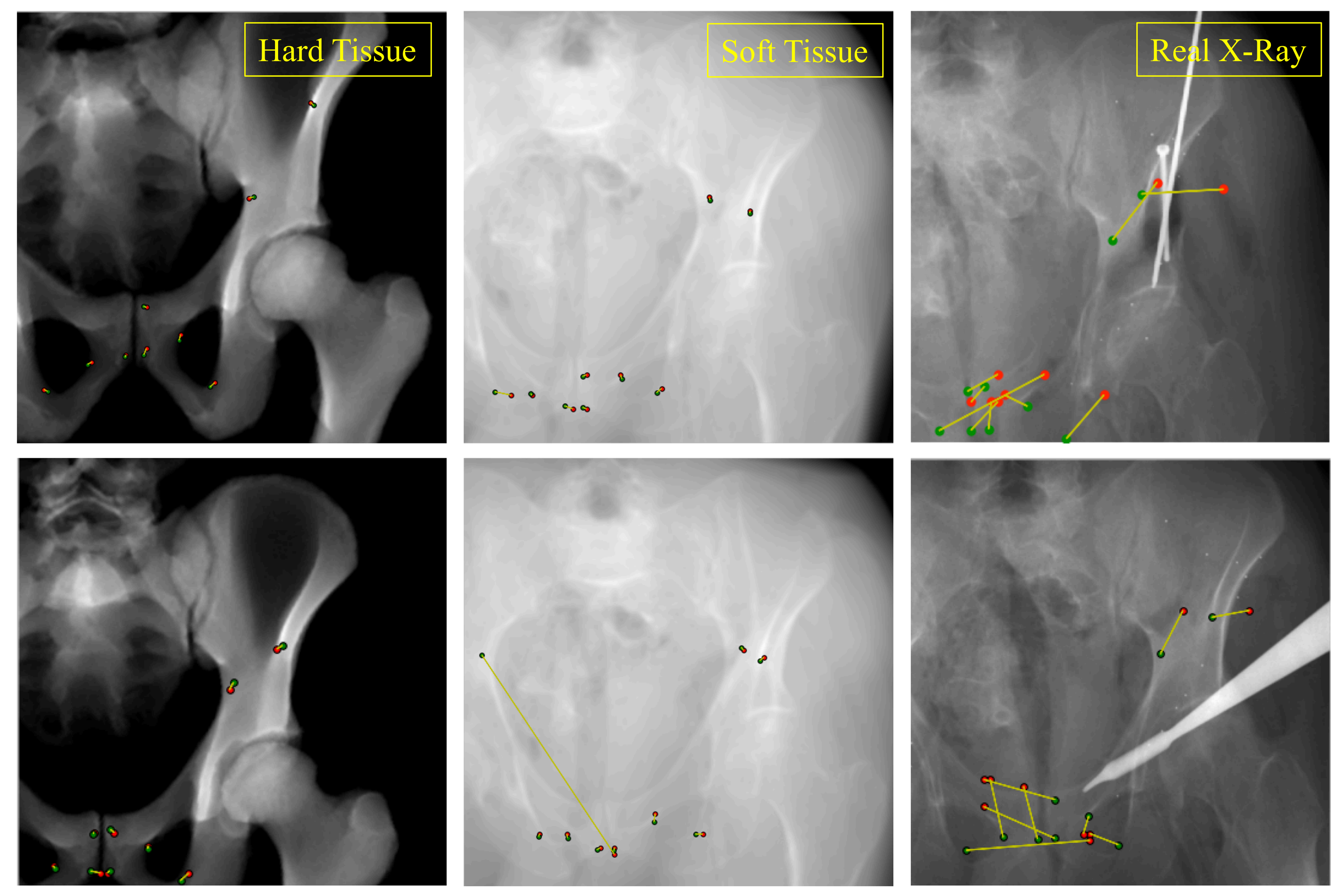


Figure 2 Landmark Detection. Red: Prediction, Green Target

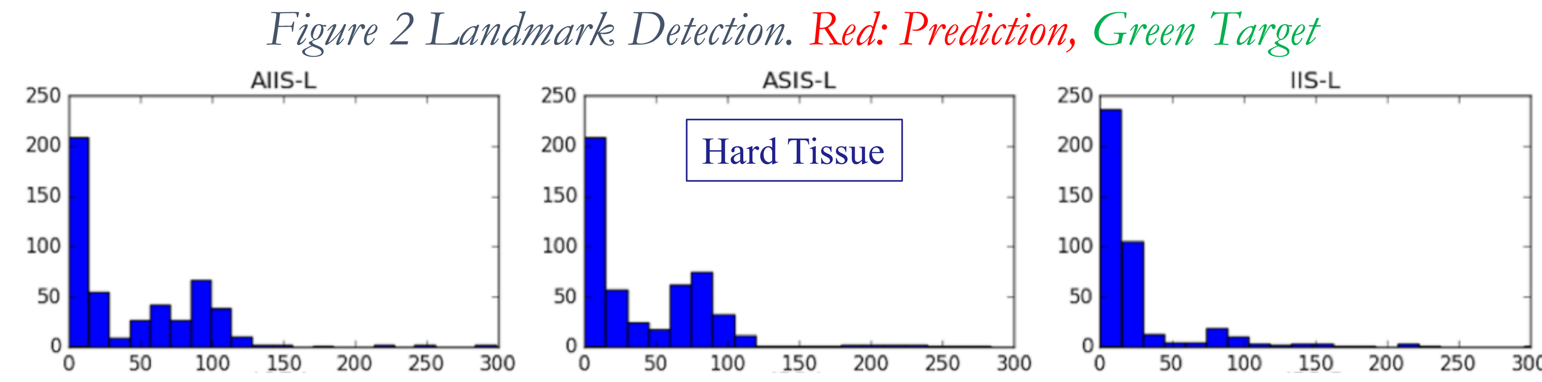


Figure 3 Selected Landmarks Prediction Error

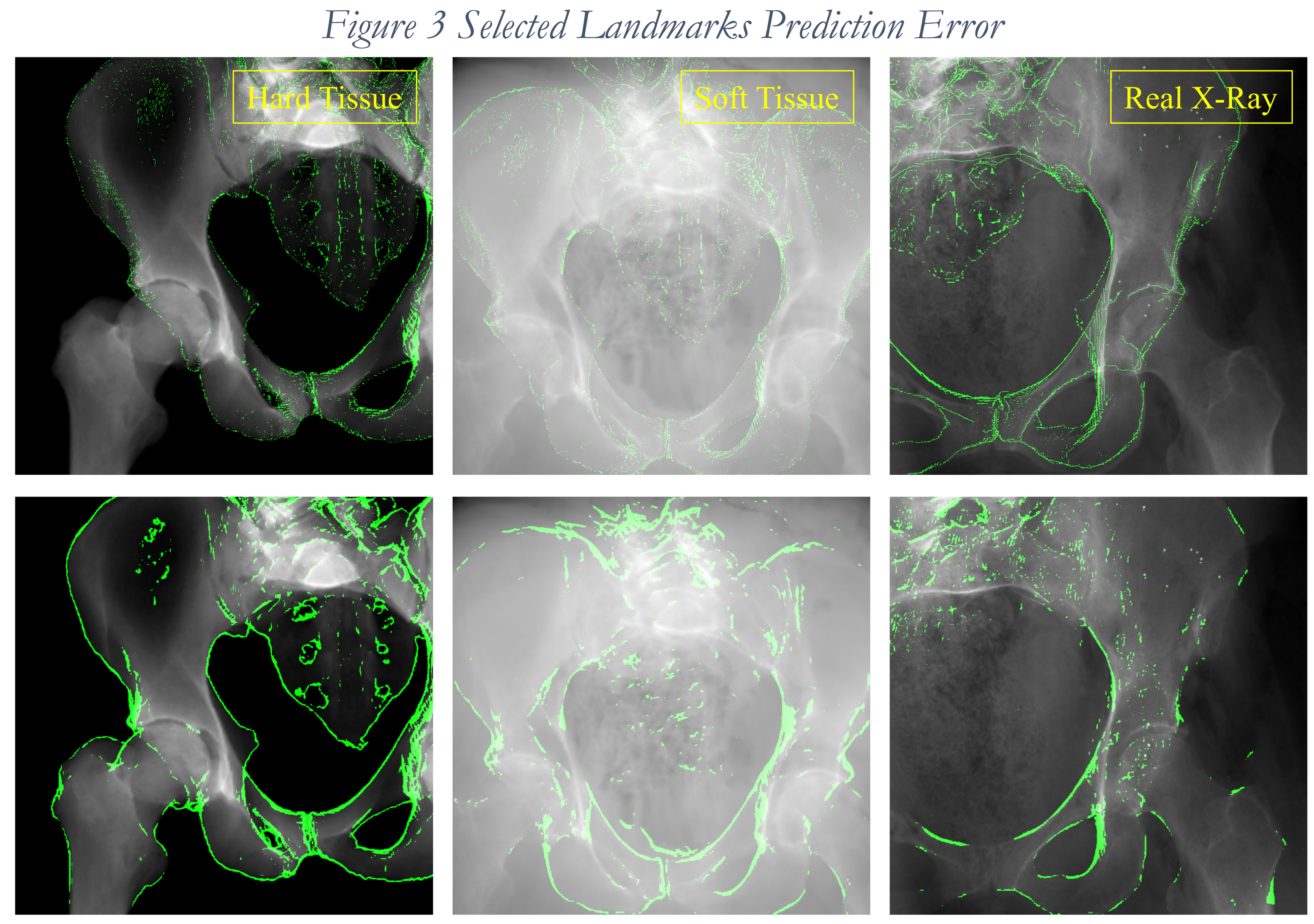


Figure 4 Contour Detection. Top: Target, Bottom: Prediction

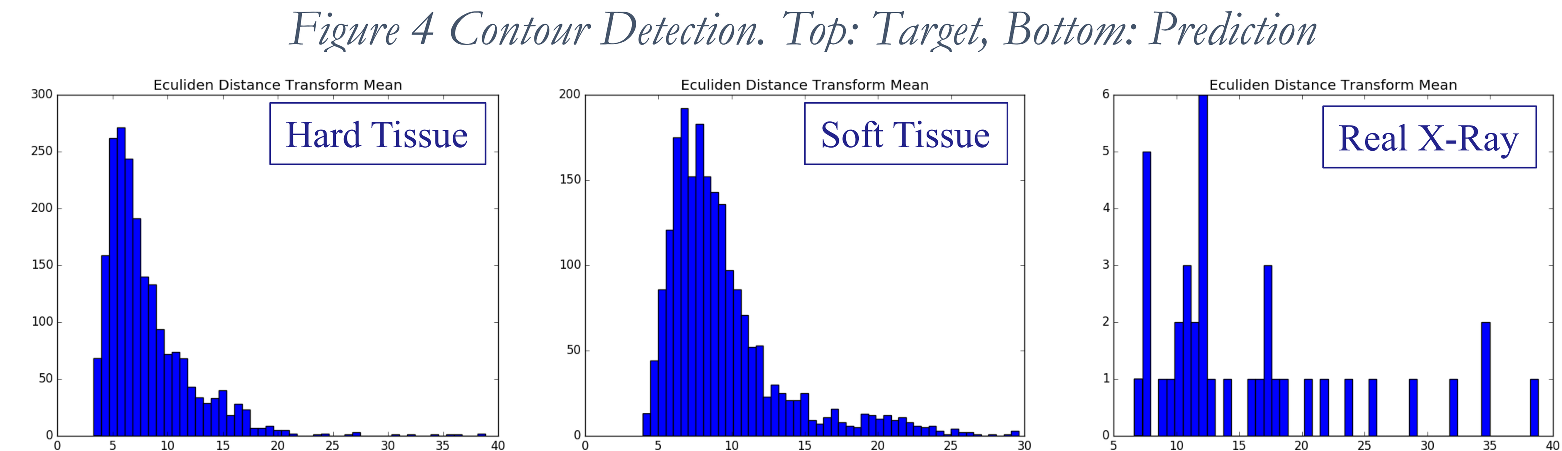


Figure 5 Edge Map Distance Transform Measure

Outcomes and Results

- **Achieved the minimum deliverables on landmark detection** by using convolutional pose machines, and performed experiments to analyze the performance and limitations.
- **Achieved the medium deliverables on tools in field of view** by introducing random region mask as data augmentation
- **Achieved the maximum deliverables on contour detection** by utilizing HED network, and successfully transferred the model trained by soft tissue simulated data to real measured X-ray data.
- **Not gathering more complex simulated data.**

Support by and Acknowledgements

- Thank Robert Grupp for weekly mentoring and data preparation.
- Thank Professor Taylor for computational resources support.