SuperLoss: A Superpixel-Guided Loss for Noisy Label Semantic Segmentation in X-Ray Images

Giannis Lazarou[1]
Jesse Krijthe[1]
Gijs van Tulder[1],
Myrthe van den Berg[1,2]

TU Delft

Introduction

Background

- Medical datasets that are used for training ML models can suffer from label noise
- Can lead to misdiagnoses and waste of time for medical professionals
- Past work shows that pixel-level information, such as superpixels, can be used to mitigate impact of noisy labels. A superpixel is a group of connected pixels that are similar in appearance.

Research Question

- Does a U-Net model with our additional loss perform better compared to one without, for different levels and fractions of noise in the training data?

Methodology

Superpixel Generation

- Consider different superpixel algorithms, aim is to overlap with ground truth mask

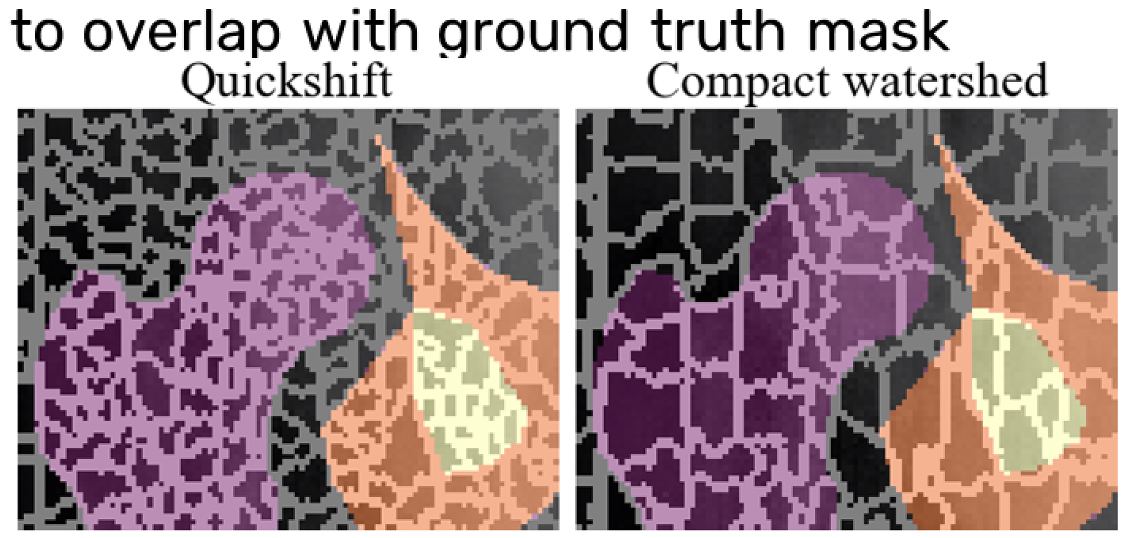

Figure 1: Comparison of different superpixel algorithms that were considered, superimposed with our ground truth mask

Noise Generation

- We inject two distinct levels of noise at different fractions (φ) in our training data
- Level 1 is a translation of a segment of the mask, emulating realistic noise
- Level 2 uses half the points needed to generate a mask, so coarser noise

SuperLoss

- We refine the model's output by majority voting semantic classes within each unique superpixel label
- Take Cross Entropy between refinement and model's output

Results

Quantitative

- We consider the Intersection-over-Union and Hausdorff Distance at 95th percentile for femur, ischium, and foramen classes. **Improvement of 1-2%** for some cases, but **sometimes performs worse**

Noise Level 1

ϕ	Model	IoU↑ Fem.	IoU↑ Isc.	IoU↑ For.	mIoU↑	HD95↓ Fem.	HD95↓ Isc.	HD95↓ For.	mHD95↓	Train
0.0	U-Net	0.946	0.899	0.897	0.915 ±0.037	**2.301**	1.630	1.484	1.805 ±1.795	4h47m
0.0	Ours	0.946	**0.905**	**0.898**	0.917±0.037	2.347	**1.587**	**1.431**	1.789 ±1.764	18h30m
0.5	U-Net	**0.939**	0.878	0.878	0.898 ±0.035	**2.667**	**1.994**	**1.602**	2.087 ±1.811	4h53m
0.5	Ours	0.928	0.878	**0.885**	0.896 ±0.040	2.855	2.043	1.717	2.205 ±2.509	18h59m
1.0	U-Net	**0.920**	0.845	0.854	0.872 ±0.039	**3.252**	2.957	1.917	2.708 ±2.766	4h57m
1.0	Ours	0.913	**0.858**	**0.865**	0.878 ±0.036	3.431	**2.315**	**1.737**	2.494 ±1.809	18h51m

Table 1: Quantitative results of our method compared to baseline for noise Level 1 at three fractions of noise (ϕ)

Qualitative

- Some boundaries seem to improve, whereas some others seem to perform worse

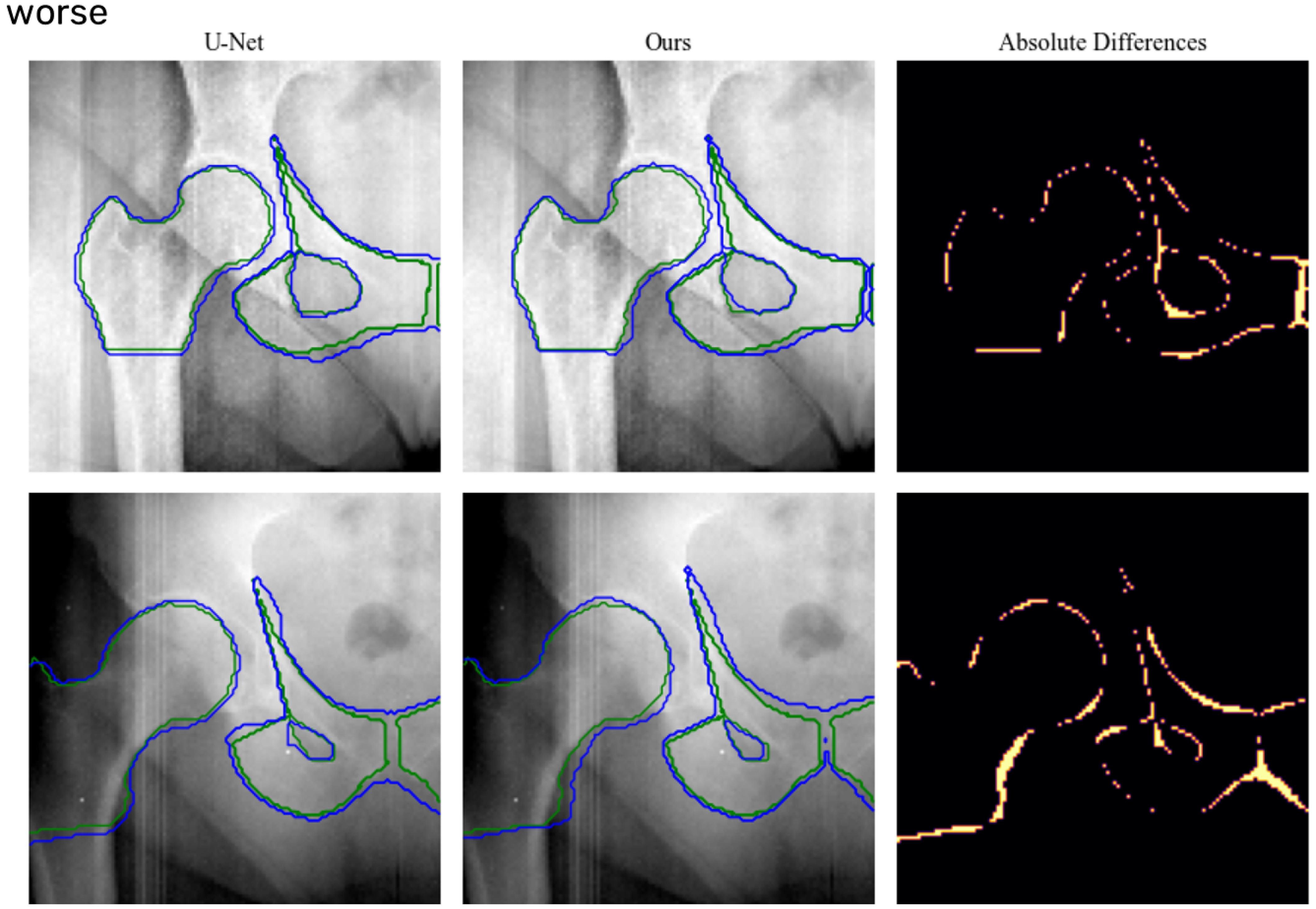

Figure 2: Qualitative results of our baseline (left) compared to our method (center), and absolute differences between them (right). Both models trained for noise Level 1 with ϕ = 1.0. Green boundaries indicate ground truth mask, and blue indicates prediction boundaries of either model

Hypothesis Testing

- Wilcoxon-signed rank test to further analyze our results
- Null Hypothesis: There is no difference in medians of our metrics. We chose our threshold to be 0.05 to reject the Null Hypothesis

Noise Level	ϕ	mIoU Statistic	mIoU p-value	mIoU $\lvert r\rvert$	mHD95 Statistic	mHD95 p-value	mHD95 $\lvert r\rvert$
1	0.0	902363	<0.001	0.167	672163	<0.001	0.068
1	0.5	1130308	<0.001	0.055	979370	<0.001	0.127
1	1.0	785833	<0.001	0.224	1015804	<0.001	0.109
2	0.5	845926	<0.001	0.196	806295	<0.001	0.199
2	1.0	844551	<0.001	0.196	994596	<0.001	0.118

Table 2: Wilcoxon-signed rank test statistic, p-value, and effect size (r) for pairwise comparisons between the baselines' and our models' mIoU and mHD95 results

- Our results are **statistically significant**, with **small to medium effect size** (*r*). Effect size quantifies the degree of differences, and has magnitude between 0 and 1

Conclusion

- Performance dependents on superpixel initialization
- In some cases, achieves small improvement in IoU and HD with small effect size
- Qualitatively does not push boundaries as much as we had expected

Limitations

- Training times are too large, due to refinement step in loss calculation
- Naïve calculation of loss, should maybe consider weighting the contribution of each superpixel separately

Affiliations

[1]EEMCS, Delft University of Technology, The Netherlands
[2]Orthopedics and Sports Medicine, Erasmus Medical Center, The Netherlands