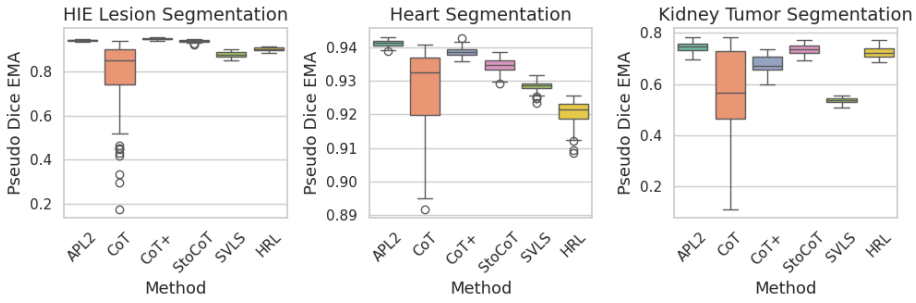
This MICCAI paper is the Open Access version, provided by the MICCAI Society. It is identical to the accepted version, except for the format and this watermark; the final published version is available on SpringerLink.

Supplementary Material: Are robust loss functions still relevant for medical image segmentation with noisy labels?

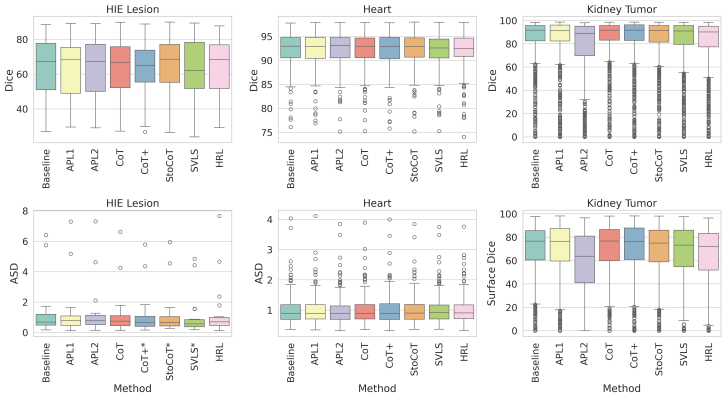
Vikram Venkatraghavan[1,2], Richard A.P. Takx[3] Niek E. van der Aa[4], Timo R. de Haan[5], Nils Planken[6,3], Manon J.N.L. Benders[4,5], and Ivana Išgum[6,1,3]

[1] Department of Biomedical Engineering & Physics, Amsterdam University Medical Center, The Netherlands.
[2] Informatics Institute, University of Amsterdam, The Netherlands.
[3] Department of Radiology & Nuclear Medicine, Amsterdam University Medical Center, The Netherlands.
[4] Department of Neonatology, Wilhelmina Children's Hospital, University Medical Center, Utrecht, The Netherlands.
[5] Department of Neonatology, Emma Children's Hospital, Amsterdam University Medical Center, The Netherlands.
[6] Department of Radiology, Mayo Clinic, Rochester, United States of America.
v.venkatraghavan@amsterdamumc.nl

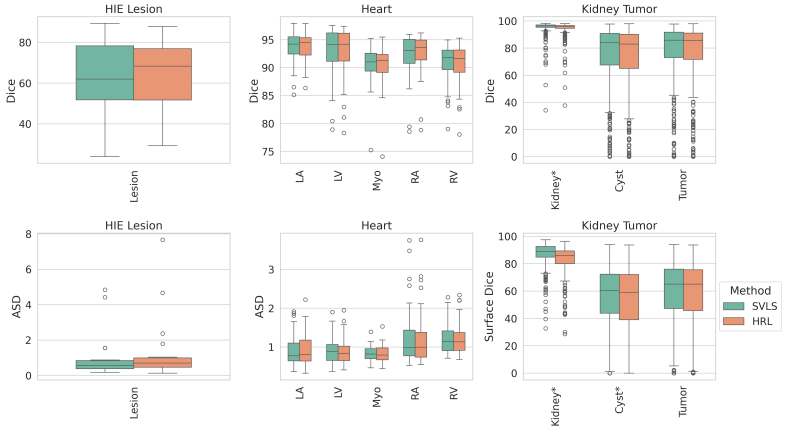
Supplementary Figures



Supplementary Fig. 1. Hyper-parameter stability: distribution of Pseudo Dice EMA values across different methods for the three medical imaging datasets, over all three cross-validation folds. Abbreviations: APL - Active-passive loss; CoT - co-teaching; StoCoT - Stochastic co-teaching; SVLS - spatially varying label smoothing; HRL - Hybrid robust loss.



Supplementary Fig. 2. Case-wise distribution of segmentation performance across all methods trained on noisy labels for different tasks. Boxplots for HIE lesion, Heart, and Kidney Tumor segmentation show the per-case Dice distribution for all methods trained on noisy labels, corresponding to the summary statistics reported in Table 1 of the main paper. Statistically significant improvement over Baseline is marked as *.



Supplementary Fig. 3. Case-wise distribution of segmentation performance comparing SVLS loss and Hybrid robust loss, corresponding to the summary statistics reported in Table 2 of the main paper. Statistically significant difference is marked as *.