Uncertainty quantification using an energy-based method applied to kidney CT images

Andreea Elena Vântu[1,2], Daniel Bunescu[1,2], Mihnea Ion[1,2], Lucian Mihai Itu[1,2]
[1]Transilvania University of Brașov; [2]Siemens SRL

Background

- Medical image analysis
- Reliable AI models
- Uncertainty quantification
- Apply post-hoc

Objective

- Use an energy-based[1] function as a proxy for identifying misclassified kidney image segmentations

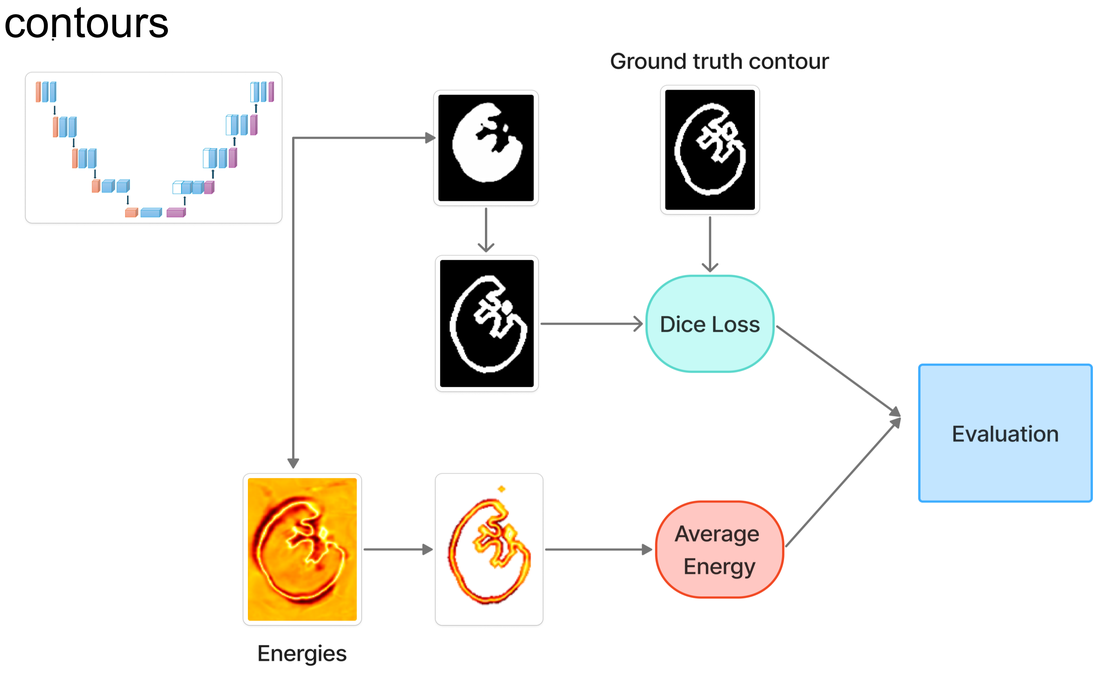
Methods

- Used KiTS21[2] and KiTS23[3] as datasets for training and testing
- K-fold Cross Validation
- 3D U-Net architecture
- Use energy function for uncertainty quantification:

$$E(x) = -T \cdot log \sum_i^K e^{f_i/T}$$

x = input
T = constant temperature
K = number of classes
$e^{f_i/T}$ = softmax partition function

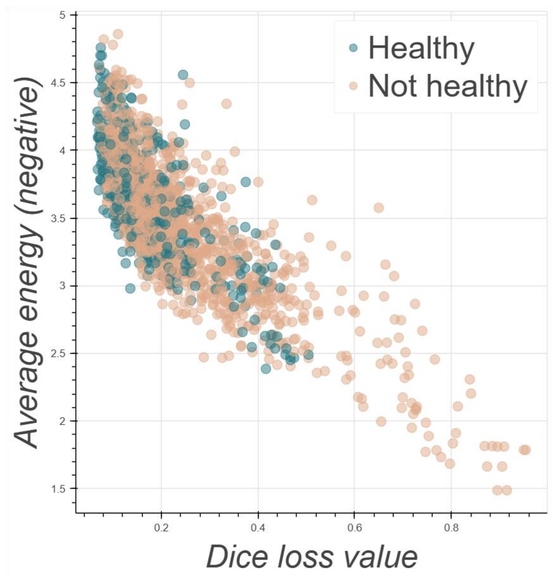
- Experiments conducted on different types of kidney contours



- Types of kidney contours:
  - inner contour
  - outer contour
  - contextual contour
- Types of contour widths:
  - increased by one voxel
  - increased by two voxels
  - increased by three voxels

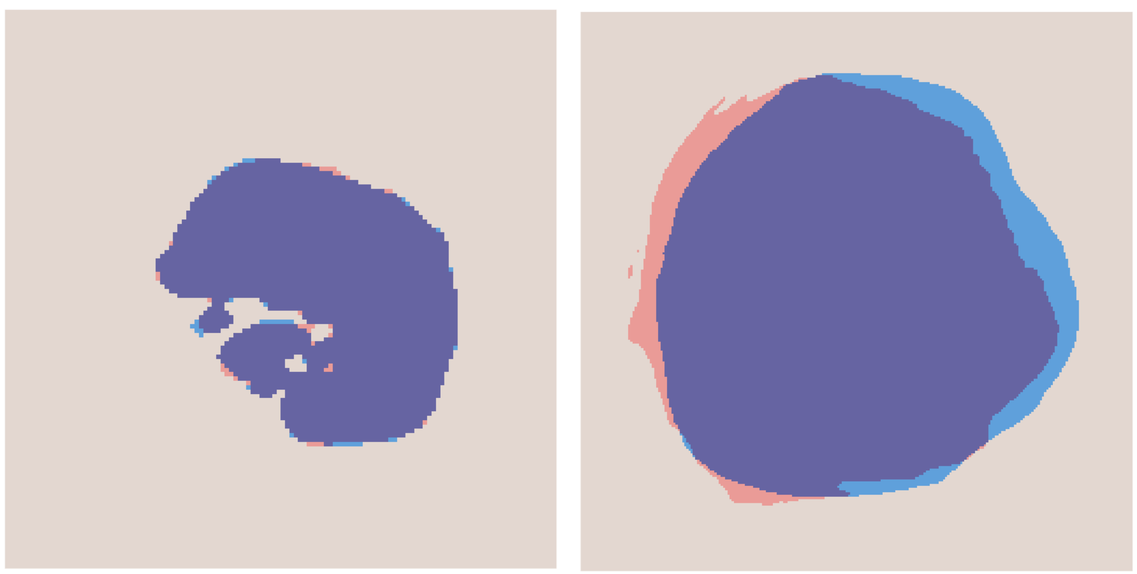
Results

According to our results, a kidney contour, with a width increased by one voxel, containing more context around the kidney yielded a better performance, having a Pearson correlation coefficient of **0.801** and a ROC-AUC score of **0.888**.



The correlation between average energy and Dice loss value evaluated on a contextual contour.

Overlapping the inner (red) and outer (blue) kidney contours. Both are placed over an energy segmentation. We refer as contextual contour the merge between these two types.



Overlapping a prediction mask (red) and a ground truth mask (blue). Left figure: the Dice loss value is 0.09 and the average energy score is -4.7. Right figure: the Dice loss value is 0.81 and the average energy score on the contour is -2.1. The Dice loss and average energy were extracted from a contextual kidney contour increased by one voxel.

Conclusion

Uncertainty quantification is an essential step in building reliable AI tools for healthcare. In this study, we determined how the energy score can be used as a proxy to identify poor kidney segmentations. A contextual kidney contour with a width increased by one voxel had better results in distinguishing such cases.

References

1. Weitang Liu, Xiaoyun Wang, John D. Owens, and Yixuan Li. 2021. Energy-basedOut-of-distribution Detection. doi:10.48550/arXiv.2010.03759 arXiv:2010.03759[cs].
2. Heller, N., Isensee, F., Maier-Hein, K.H., et al. (2021). *The state of the art in kidney and kidney tumor segmentation in contrast-enhanced CT imaging: Results of the KiTS19 challenge*. *Med Image Anal*, 67, 101821. https://doi.org/10.1016/j.media.2020.101821
3. Heller, N., Isensee, F., Trofimova, D., et al. (2023). *The KiTS21 Challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase CT*. arXiv:2307.01984. https://doi.org/10.48550/arXiv.2307.01984

12th ACM Celebration of Women in Computing: womENcourage™
Brașov, Romania
17-19 September, 2025
Theme: Computer Science: a Catalyst for Educational Change