2026 JOINT AAPM | COMP MEETING
JULY 19–22 | VANCOUVER, BC

A Smart Calibrated Probabilistic U-Net for Aleatoric and Epistemic Uncertainty Quantification in Medical Image Segmentation

Allen J. Chao[1]
[1]University of California, Irvine, Irvine, CA

Purpose

Image segmentation is a critical yet challenging step in the radiotherapy workflow. Although many AI-based segmentation models report high accuracy, substantial epistemic and aleatoric uncertainties persist. This work proposes a probabilistic U-Net framework to explicitly quantify and calibrate segmentation uncertainty, with the goal of improving reliability for clinical decision-making.

Methods & Materials

We developed a Smart Calibrated Probabilistic U-Net (SCP-UNet) that departs from conventional probabilistic U-Net designs that employ separate prior and posterior encoders (Fig. 1). The proposed model employs a single encoder with a fixed standard normal prior *N*(0, 1), eliminating mask-dependent encoding. Bottleneck features are mapped to latent variables through adaptive pooling and linear projections, followed by reparameterized sampling to generate diverse predictions. The model was optimized using a variational autoencoder (VAE)-style ELBO (Evidence Lower Bound) objective combining binary cross-entropy, Dice, Kullback-Leibler (KL) divergence, and calibration losses. SCP-UNet was trained using AdamW optimizer with mixed-precision optimization and evaluated separately on prostate (4 structures, 42 cases) and head-and-neck (5 structures, 35 cases) datasets using segmentation accuracy, uncertainty metrics, and uncertainty-error correlation.

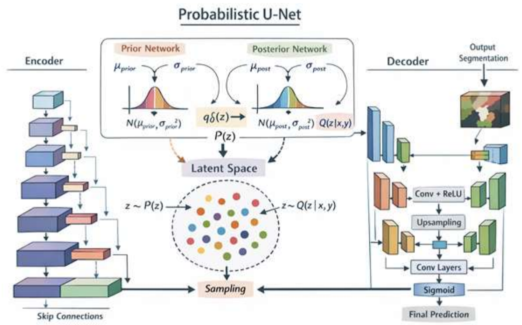

Figure 1. The architecture of the probabilistic U-Net for uncertainty estimate.

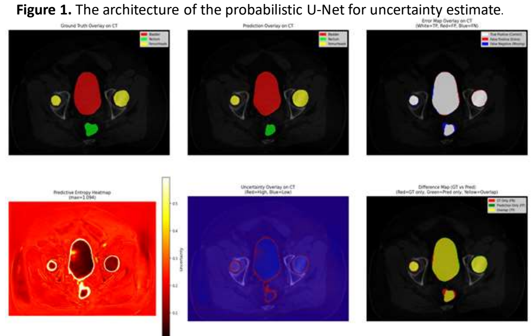

Figure 3. An illustration of a prostate case. The top panel demonstrates the segmentation (ground truth, prediction, and difference map) overlaid on CT.. The bottom shows the predictive entropy heat map and uncertainty map.

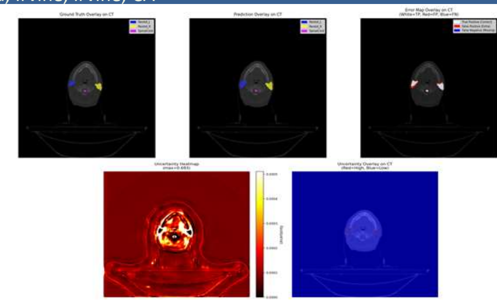

Figure 2. An illustration of a head and neck case. The top panel demonstrates the segmentation (ground truth, prediction, and difference map) overlaid on CT.. The bottom shows the predictive entropy heat map and uncertainty map.

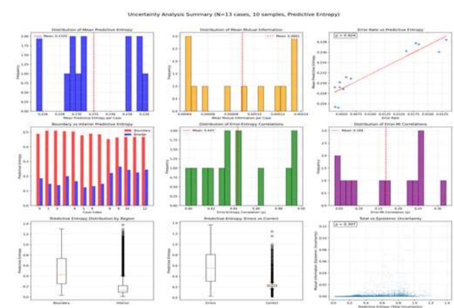

Figure 4. Uncertainty analysis for the test case of a prostate patient.

Results

Figures 2 and 3 illustrate the sementation, predictive entropy heat map, and uncertainty map for a prostate and head-and-neck case, respectively. Figure 4 illustrate the uncertainty analysis for the prostate case in Fig. 3.

On the test (validation) sets, mean Dice/HD95 were 0.8556 ± 0.1932/8.77 ± 1.43mm (0.7959 ± 0.1127/8.56 ± 1.12mm) for prostate and 0.7034 ± 0.2483/1.69 ± 0.84mm (0.7246 ± 0.1037/1.78 ±0.56mm) for head-and-neck, respectively. Mean predictive entropy was substantially higher for prostate than for head-and-neck (0.2315 vs 0.00046), corresponding to overall error rates of 0.96% and 29.09%, respectively. Moderate segmentation error-entropy correlations were observed for both sites (0.4484 prostate; 0.3855 head-and-neck). Epistemic uncertainty ratios were 0.0 for prostate and 0.086 for head-and-neck.

Conclusions

The proposed SCP-UNet achieved well-calibrated aleatoric uncertainty for prostate segmentation, whereas head-and-neck segmentation exhibited overconfidence and elevated error rates, indicating suboptimal uncertainty calibration. These findings highlight the dataset- and task-dependent nature of uncertainty quantification. This probabilistic segmentation framework enables explainable uncertainty estimation to support informed segmentation review and quality assurance.