Supplementary : Uncertainty Guidance for Medical Image-to-Image Translation

Anuja Vats, Ivar Farup, Marius Pedersen, Kiran Raja
Department of Computer Science, NTNU, Gjøvik, Norway

Additional Results on CPC-Paired Dataset

Table 1 shows the performance of image translation from Normal to FICE mode on the CPC-paired dataset, with and without UAR. We measure SSIM, PSNR, LPIPS and RRMSE metrics under the impact of three types of noise, Gaussian, Uniform and Implulse as before. As before, It can be seen that in general UAR improves the reconstruction quality. Figure 2. shows uncertainty maps from this dataset. Note the high SSIM values associated with visibly poor reconstructions (rows 2 and 3 in Figure 2). This discrepancy occurs because SSIM relies on statistical properties rather than perceptual quality. Therefore, LPIPS and RRMSE provide more reliable measures of reconstruction quality. These are also where UAR has the most significant improvement across various noise types and levels.

		Approach	SSIM ↑	PSNR ↑	LPIPS ↓	RRMSE ↓
Gaussian	$\mathcal{N}(0, 0.001)$	Baseline	0.781	33.674	0.507	0.3340
		UAR (Ours)	**0.902**	**36.223**	**0.441**	**0.212**
Uniform	$\mathcal{U}(0, 0.1)$	Baseline	0.862	**35.055**	0.468	0.297
		UAR (Ours)	**0.864**	34.695	**0.418**	**0.257**
Impulse	$\mathcal{I}(p = 0.005)$	Baseline	0.744	**32.288**	0.570	0.432
		UAR (Ours)	**0.758**	30.321	**0.557**	**0.368**

Table 1. Impact of uncertainty-guidance on reconstruction quality on CPC-Paired dataset

Ablation I: Other variation based losses

This section compares qualitatively the uncertainty maps generated from the different variation-based penalties discussed in Sec. 5 of the main paper. As seen in Fig. **??**, all variants with regularization show consistently sparser uncertainity maps compared to the baseline, while the regions of uncertainty are consistent. The UAR and UAR_{aniso} generate sharper uncertainty maps with preference for edges, as compared to the UAR_{L2} which produces smoother looking uncertainty maps.

The optimal choice of regularization penalty may vary depending on the specific application and the purpose of the uncertainty estimation. For instance, in a medical context, selecting the appropriate penalty could involve consulting with medical professionals to determine which uncertainty maps are most meaningful to them. This collaborative approach could also inform the development of new penalties that incorporate doctors' insights as priors for uncertainty estimation.

Ablation II : λ Range

We conducted experiments with extreme values of λ to assess the sensitivity and optimal range of λ values, specifically testing λ at 10^{-12}, 10^{-7}, and 10^{-4}. As shown in Table 2, a high regularization weight of $\lambda = 10^{-4}$ imposes an excessively strong regularization effect on β, leading to suboptimal reconstruction performance. Fig. 3 illustrates that this causes the predicted β values to become overly similar, suppressing potentially meaningful disparities.

In contrast, $\lambda = 10^{-7}$ achieves a balance, providing effective regularization without excessively homogenizing the β values. This balance results in improved reconstruction performance, as seen by superior SSIM, PSNR, and LPIPS values compared to the baseline, as detailed in Table 2. Using $\lambda = 10^{-12}$ reflects a more cautious approach, yielding good results while maintaining robustness. We anticipate that all values within this range will be beneficial for both uncertainty estimation and reconstruction tasks. The optimal value for λ likely resides in the range $[10^{-7}, 10^{-12}]$, with a preference towards the vicinity of 10^{-12}.

Anonymized Code repository:
https://anonymous.4open.science/r/Uncertainity-Aware-Regularizatopn–UAR–ECF7/README.md

Discussion of potential negative societal impact

Environmental Impact. Training deep learning models, such as the U-Net based model employed in this paper, has environmental implications due to the high computational power required. In our attempts to reduce this impact, we conducted parameter search for λ not across all possi-

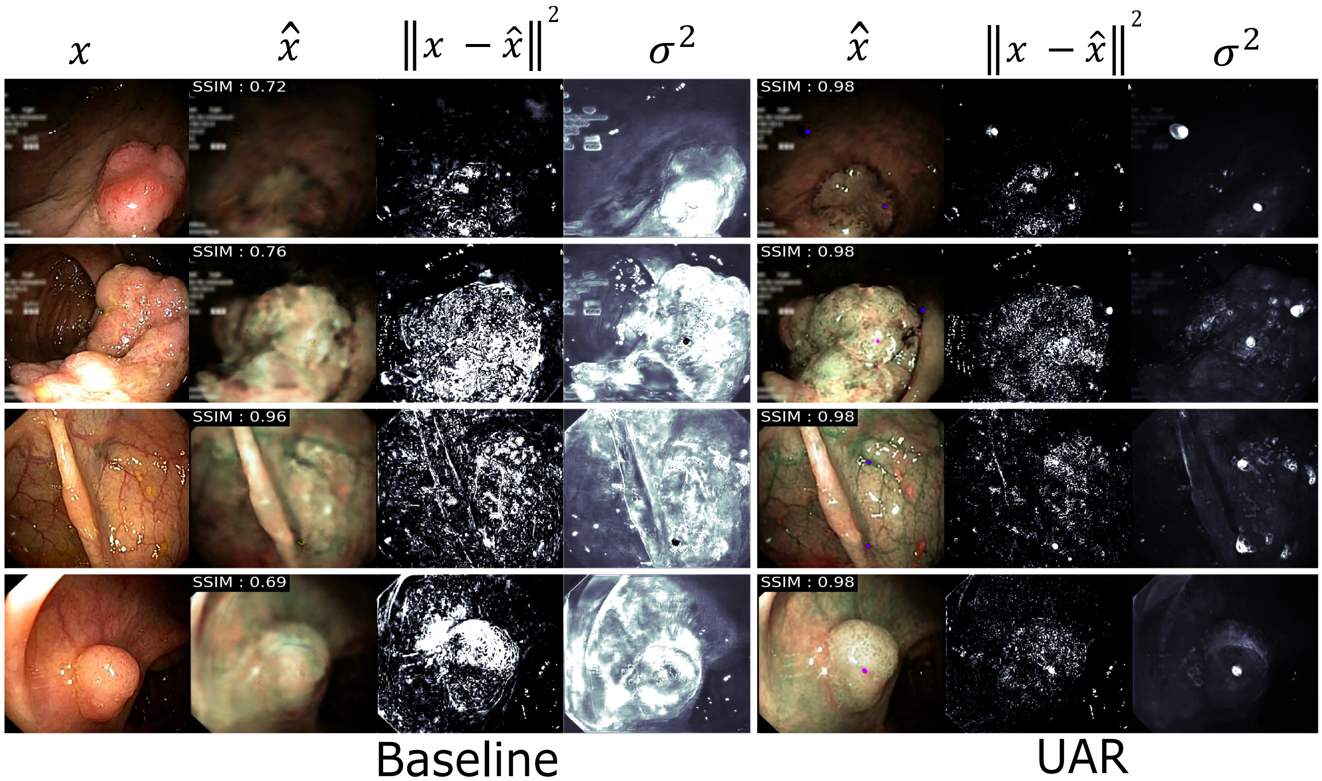

Figure 1. Qualitative comparisons of aleatoric uncertainty maps σ^2 for CPC-paired dataset.

Model	SSIM↑	PSNR↑	LPIPS↓	RRMSE↓
Baseline	0.925	28.714	0.128	0.174
$\lambda = 10^{-4}$	0.916	26.785	0.131	**0.166**
$\lambda = 10^{-7}$	**0.928**	**30.184**	0.129	0.252
$\lambda = 10^{-12}$	**0.931**	**29.340**	**0.126**	**0.148**

Table 2. Comparison of reconstruction performance across varying values of the regularization weight λ. Bold values denote metrics that surpass the baseline performance.

ble values, but rather at key points within the range. This approach allows us to identify not the exact optimal value, but at least an interval where the model performs favorably. By doing so, we significantly reduce the computational resources and environmental costs associated with parameter optimization.

Vulnerabilities and Adversarial Attacks. Quantifying uncertainty in medical image translation provides insights into the areas where the network is unsure of its predictions. While this information is crucial for improving model robustness and reliability, it also exposes potential vulnerabilities. Malicious actors could exploit these uncertainties to design adversarial attacks, specifically targeting the regions or classes where the model exhibits higher uncertainty. Such attacks could compromise the integrity of medical diagnoses and treatments, posing serious risks to patient health. Further, adversaries could also intentionally inject biases or exploit the blind spots correlated with high uncertainty, leading to biased or inaccurate outcomes.

Figure 2. Qualitative comparison of different variation-based penalties on uncertainty maps.

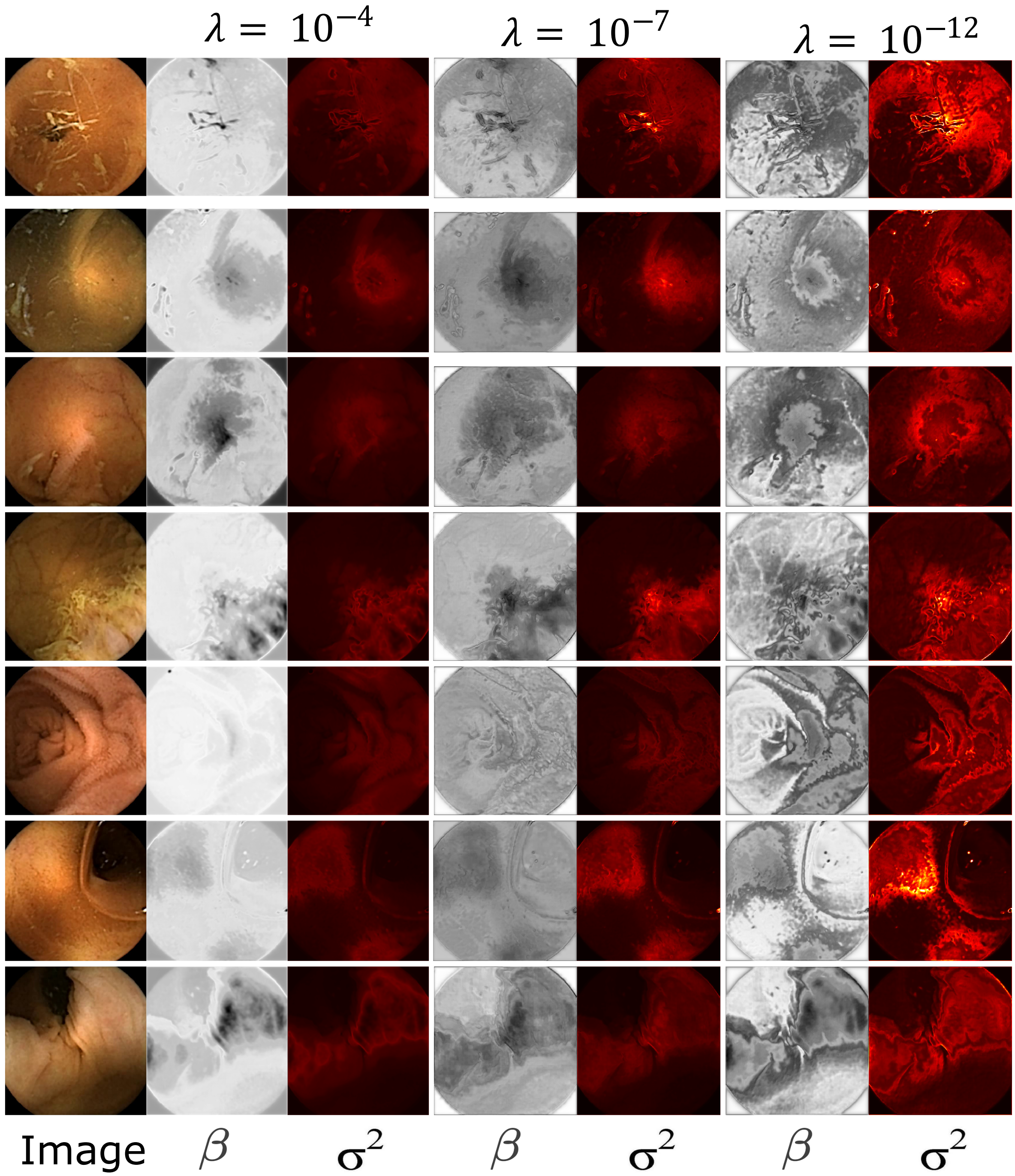

Figure 3. Qualitative comparisons of β and the aleatoric uncertainty maps σ^2 for different λ values. Observe that although regions of high uncertainty remain consistent, the influence of a large λ manifests in the relative differences between β values across image regions. For high λ values, β values appear more uniform, whereas lower λ values allow disparities in shapes β for the per-pixel residual distribution while reducing variations, resulting in significantly enhanced contrast in the uncertainty maps. $\lambda = 10^{-12}$ uses color scaling to enhance visual clarity.