Supplementary Material
Harnessing Uncertainty in Domain Adaptation for MRI Prostate Lesion Segmentation

Eleni Chiou[1,2], Francesco Giganti[3,4], Shonit Punwani[5], Iasonas Kokkinos[2], and Eleftheria Panagiotaki[1,2]

[1] Centre for Medical Image Computing, UCL, London, UK
[2] Department of Computer Science, UCL, London, UK
[3] Department of Radiology, UCLH NHS Foundation Trust, London, UK
[4] Centre for Medical Imaging, Division of Medicine, UCL, London, UK
eleni.chiou.17@ucl.ac.uk

1 Qualitative results

In this section we provide further qualitative results that could not be presented in the paper due to space constraints. Figure 1 shows the mapping from mp-MRI to VERDICT. Our approach is able to generate multiple outputs while preserving the critical structure corresponding to the prostate lesion.

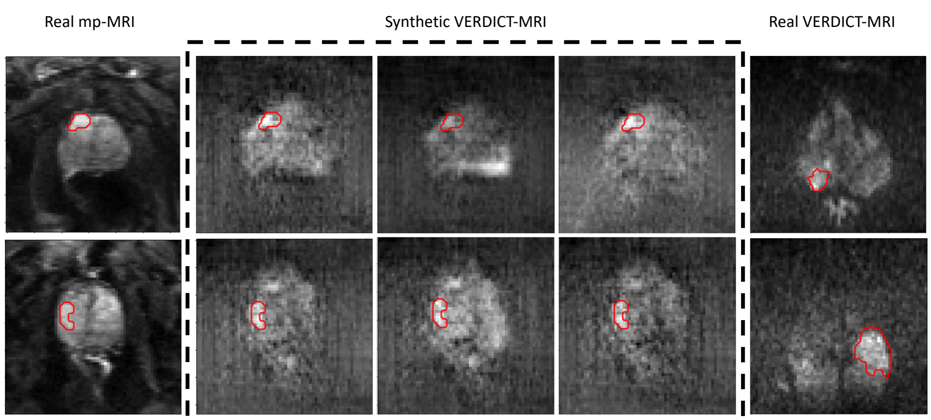

Fig. 1. One-to-many mapping from one mp-MRI (left) to three VERDICT-MRI translations (middle) for two different patients (rows): Our network can generate samples with both local and global structure variation, while at the same time preserving the critical structure corresponding to the prostate lesion. The right column shows two real VERDICT-MRI samples as an example of data from the target domain.

In Figure 2 we present lesion segmentation results produced by the different models for two patients. i) MUNIT + $\mathcal{L}_{Seg}^{Synth}$ + RAs (Ours): we use MUNIT and segmentation supervision for the translation and introduce RAs in the segmentation network. ii) CycleGan + $\mathcal{L}_{Seg}^{Synth}$ + RAs: we use CycleGan and segmentation supervision to perform the translation, and introduce residual adapters (RAs) in the segmentation network. iii) VERDICT-MRI only: we train the segmentation network only on real VERDICT-MRI.

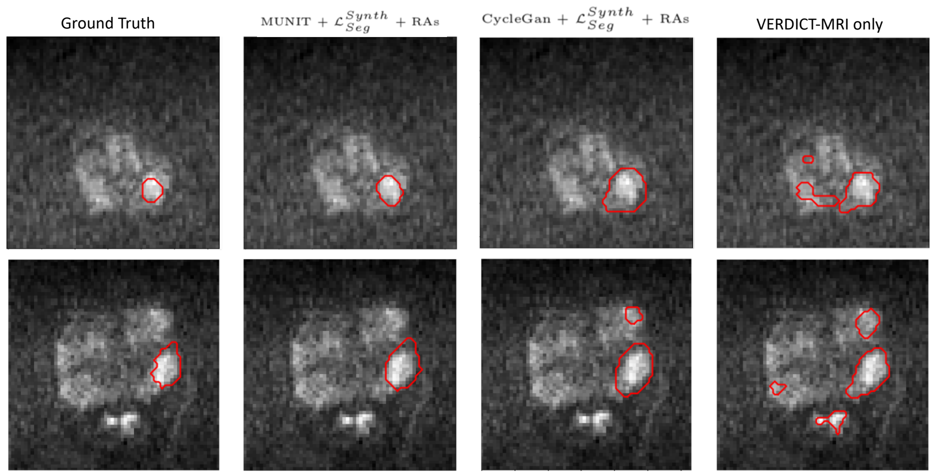

Fig. 2. Lesion segmentation results for two patients. MUNIT + $\mathcal{L}_{Seg}^{Synth}$ + RAs (Ours): we use MUNIT and segmentation supervision for the translation and introduce RAs in the segmentation network. Cyclegan + $\mathcal{L}_{Seg}^{Synth}$ + RAs: we use CycleGan and segmentation supervision for the translation and introduce RAs in the segmentation network. VERDICT-MRI only: Trained only on real data.

2 Quantitative results

In this section we provide further quantitative results that could not be presented in the paper due to space constraints. To experimentally validate that sampling different style codes enhances the performance, we perform two experiments: i) we keep the style code of the target domain fixed during the translation and ii) we use the encoded style code of the source domain. We evaluate the performance based on the mean recall, precision, dice similarity coefficient (DSC), and average precision (AP) across 5 folds. The results (Table 1) show that indeed sampling different style codes improves the performance. Figure 3 shows the impact of residual adapters in the performance for different dataset sizes. We vary the percentage of real data while keeping fixed the amount of synthesized data.

Introducing residual adapters in the segmentation network while using MUNIT and segmentation supervision (MUNIT + $\mathcal{L}_{Seg}^{Synth}$ + RAs) during the translation systematically improves performance for different dataset sizes.

Table 1. Impact of sampling on the performance. Average recall, precision, dice similarity coefficient (DSC), and average precision (AP) across 5 folds. The results are given in mean (±std) format.

Model	Recall	Precision	DSC	AP
MUNIT + $\mathcal{L}_{Seg}^{Synth}$ + fixed s_T	68.4 (±9.0)	65.3 (±7.9)	67.0 (±8.8)	70.0 (±8.9)
MUNIT + $\mathcal{L}_{Seg}^{Synth}$ + encoded s_S	61.5 (±12.4)	68.5 (±9.2)	64.5 (±10.6)	67.4 (±10.6)
MUNIT + $\mathcal{L}_{Seg}^{Synth}$ (Ours)	**71.8** (±7.8)	**68.0** (±6.8)	**69.8** (±7.9)	**73.5** (±8.1)

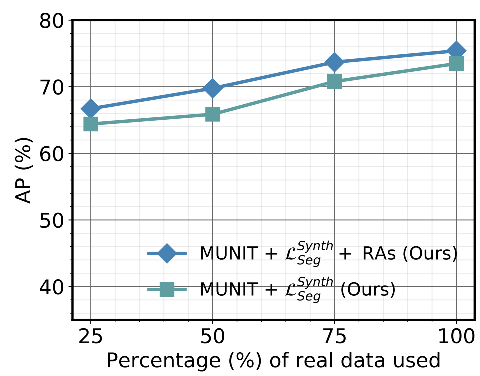

Fig. 3. Impact of residual adapters (RAs) in the average precision (AP) for different dataset sizes. We vary the percentage of real data while keeping fixed the amount of synthesized data. Introducing residual adapters in the segmentation network while using MUNIT and segmentation supervision (MUNIT + $\mathcal{L}_{Seg}^{Synth}$ + RAs) during the translation systematically improves performance for different dataset sizes.