

## Limited Paired Labels Suffice for Binary but Not Multiclass Ultrasound Segmentation

Precious Kolawole, Matthew S. Holden

Carleton University, School of Computer Science, Ottawa, Canada

**Introduction:** Ultrasound imaging provides real-time, radiation-free views of internal anatomy and is widely used for diagnosis and image-guided procedures, but the raw images are noisy and difficult for both clinicians and automated systems to interpret. One way to make ultrasound easier to use is to simplify the images through segmentation, where the relevant anatomy is separated from the background. However, traditional fully supervised segmentation methods require many scans with detailed annotated masks, which is expensive and especially challenging in resource-limited settings. Recent unpaired image translation and semi-supervised methods promise to reduce the need for annotation. This motivates our study of how effectively ultrasound images can be simplified through segmentation when only limited or no paired annotations are available. Our objectives were to: (1) determine whether unpaired domain translation can produce clinically useful ultrasound segmentations; and (2) examine how task complexity (binary vs. multiclass) changes these requirements.

**Methods:** We used the Open Kidney Ultrasound Dataset (534 abdominal B-mode images with expert annotations for four kidney regions: cortex, capsule, medulla, and central echo complex) and compared three learning regimes. A fully supervised baseline used U-Net++ (ResNet-101 encoders pretrained on ImageNet) trained on all paired image-mask samples for both binary and multiclass tasks. An unpaired CycleGAN ultrasound-mask model with U-Net generators and PatchGAN discriminators was trained on non corresponding image-mask pairs, with an additional identity loss for the binary experiments. A semi-supervised hybrid retained the CycleGAN adversarial and cycle-consistency losses on all unpaired data while also applying a cross-entropy segmentation loss to a small set of paired image-mask examples (with identity loss for the binary case). Images were center-cropped around the kidney, resized to  $256 \times 256$  pixels, normalized, and augmented with random flips and rotations. To prevent information leakage, we split the dataset into two non-overlapping halves (Set A and Set B, 267 pairs each) and within each used a 70/15/15 train/validation/test split. Models were implemented in PyTorch and trained for 40 epochs with Adam optimizer ( $\text{lr } 1 \times 10^{-4}$ ). CycleGAN models used images from Set A and masks from Set B to ensure a truly unpaired regime. Evaluation used paired images and masks from the Set A validation and test sets, and performance was reported using Dice and Intersection-over-Union (IoU) scores.

**Results:** Fully supervised models achieved the best performance, with Dice/IoU of 0.900/0.820 for binary kidney vs. background and 0.627/0.470 for multiclass segmentation using all paired data. Unpaired CycleGAN performed very poorly in both settings (binary Dice/IoU 0.058/0.030; multiclass 0.043/0.024), indicating that purely unpaired domain translation was not sufficient for proper US simplification. Semi-supervised hybrid models, at  $N = 93$  paired samples, reached a binary Dice of 0.842 and a multiclass Dice of 0.274; this improved over the unpaired CycleGAN but fell short of fully supervised multiclass performance. Across methods and data regimes, binary consistently outperformed multiclass and achieved higher Dice scores with fewer paired examples.

**Conclusions:** Fully supervised models achieved ultrasound simplification for binary tasks (Dice  $> 0.7$ ), while multiclass segmentation requires substantially more annotated data. Unpaired CycleGAN models performed poorly, and semi-supervised hybrids, although better than unpaired training, did not match fully supervised multiclass performance, indicating that unpaired images and masks alone have limited value. We therefore recommend using paired datasets whenever possible and using semi-supervised methods only to extend limited labels. For future work, paired data could be collected with tracked ultrasound. Tracked ultrasound can be used to reconstruct a 3D ultrasound volume that is registered to a CT or MR image (or to fiducials visible in both). Segmentation can then be performed once on the CT/MR image, and the simplified 3D anatomy can be mapped back to any ultrasound slice using the registration transform. This could support a wide range of ultrasound simplification tasks.