

## Latent Feature-Enhanced Residual UNET for Accurate Segmentation under Low Annotation Constraints

Negar Mehrabi<sup>1</sup> and Geoffrey Gale Handsfield<sup>2,3</sup>

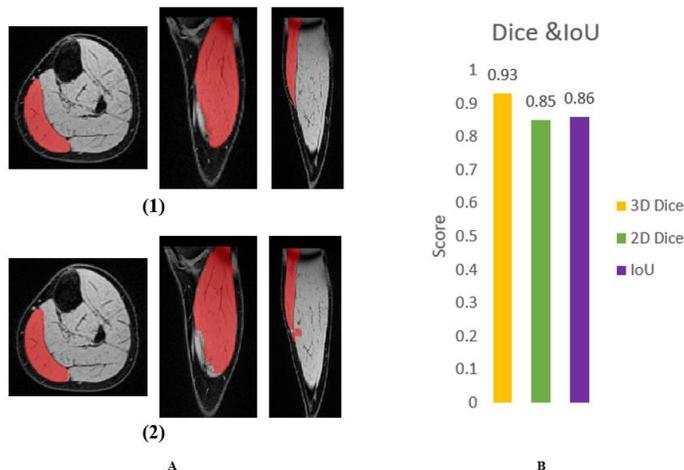
<sup>1</sup> Department of Applied Physical Sciences, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America

<sup>2</sup> Department of Orthopedics, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States of America

<sup>3</sup> Lampe Joint Department of Biomedical Engineering, University of North Carolina at Chapel Hill and NC State University

**Disclosures:** Authors NM and GGH have no financial or other interests to disclose. mehrabi@email.unc.edu

**Introduction:** Quantitative analysis of musculoskeletal MRI plays an important role in orthopedic diagnosis and treatment planning, where precise segmentation of the volume of interest (VOI) is an essential prerequisite. Manual segmentation is time-consuming, requires expert input, and is prone to inter- and intra-observer variability. These challenges become even more prohibitive for large datasets, as manual segmentation time scales directly with dataset size. Deep learning offers a powerful solution to this problem, but its efficiency highly depends on the availability of large, annotated datasets for network training. This dependency limits its applicability in scenarios where prior labeled data is scarce. To address this limitation and reduce the annotation burden, we propose a deep learning framework capable of robust, accurate segmentation trained on a single patient's lower-limb MRI volume using only 10 manually segmented slices of the medial gastrocnemius muscle as ground truth. Our network architecture drastically reduces segmentation time and reliance on extensive labeled datasets, highlighting its potential for clinical translation in settings where rapid, reliable diagnosis is required. **Methods:** We developed our model using a single volumetric patient MRI scan of the lower limb, from which only 10 manually segmented slices of the healthy medial gastrocnemius muscle were available as ground truth. To increase the utility and robustness of this highly limited ground truth data set, we implemented a Residual UNet architecture integrated with a ResNet-inspired latent encoder. The input to the segmentation network consists of normalized five-slice stacks extracted along the anatomical axis of the MRI volume. This stacking strategy ensures that the network receives 3D contextual information from neighboring slices, thereby compensating for the sparse manual labeling. The latent encoder compresses each multi-slice input into a compact embedding vector that encodes global structural information. These embeddings are then expanded and concatenated with the UNet feature maps, enriching the local convolutional features with global anatomical priors. To increase the generalization of the training, we introduced extensive augmentation of the few available masks, including morphological dilation/erosion and stochastic noise injection, ensuring robustness to boundary variations and image perturbations. A composite loss function was designed, combining Dice, binary cross-entropy, boundary-aware loss, Tversky, and a small-object penalty, with the objective of reducing false positives and constraining over-prediction of muscle boundaries. From the 10 manual segmentations, 8 were used for training (80%) and 2 for validation (20%). To prevent overfitting given the limited data, early stopping was employed, halting training once the validation loss plateaued. This strategy, combined with the latent feature augmentation, enabled effective model convergence from only a single patient MRI volume with 10 manually segmented medial gastrocnemius masks as ground truth, while still preserving strong generalization capacity across unseen data. **Results:** The proposed model demonstrated high segmentation accuracy and stability across the entire medial gastrocnemius muscle volume despite the limited ground truth masks. Quantitative evaluation yielded a mean 2D Dice coefficient of 0.88 and a 3D Dice coefficient of 0.93, alongside an Intersection-over-Union (IoU) score of 0.86, indicating excellent spatial overlap between predicted and ground truth masks. The model generalized to unlabeled MRI slices, producing smooth, anatomically coherent segmentation boundaries. The integration of latent embedding, multi-slice contextual inputs, and advanced residual blocks significantly enhanced the extraction of features under low-data constraints. **Conclusion and future outlook:** These findings validate the proposed Latent Feature-Enhanced ResNet architecture and training strategy as an effective solution for accurate muscle segmentation with minimal manual labeling. The framework enables rapid and reliable volume-of-interest delineation, supporting more accurate diagnosis and treatment in time-constrained clinical settings. This capability holds broad application in both research and clinical workflows, including personalized muscle modeling and other applications where fast, precise segmentation is essential.



**A)** Axial, coronal, and sagittal MRI views of the lower limb, highlighting manual segmentations (1) and model-predicted segmentations (2) of the medial gastrocnemius (MG) muscle. **B)** Segmentation performance metrics.