DOI 10.5162/Ultrasonic2025/E4-a1

RG-UNet: Automated Muscle Ultrasound Segmentation Using the U-Net Architecture Augmented with Monogenic Phase Asymmetry Maps

Pascal Engl[1], *Hossein S. Aghamiry*[2], *Johannes Schwerdt*[1], *and Klaus-Vitold Jenderka*[1]
[1] *University of Applied Sciences Merseburg, Eberhard-Leibnitz-Str. 2, 06217 Merseburg, Germany*
[2] *Department of Radiology, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany*
pascal.engl@hs-merseburg.de

Abstract: This study presents a novel deep learning approach for automated image segmentation of dynamic ultrasound images of the *vastus lateralis* muscle, which addresses key challenges like data availability and image variability. The proposed RG-UNet augments a standard lightweight U-Net architecture with a secondary input channel derived from monogenic phase asymmetry analysis, providing the network with intensity-invariant structural information for tissue boundaries and muscle fascicles and by this achieving superior segmentation performance.

Keywords: automated image segmentation, U-Net architecture, musculoskeletal ultrasound, monogenic signal, deep learning

Introduction

In medical diagnostics, ultrasound applications often require the identification and characterization of specific regions of interest (ROIs) within B-mode images. To effectively analyze large sets of images and their corresponding ROIs in real-time, automated image segmentation (IS) is essential. IS not only facilitates the detection of abnormalities but also allows for a systematic quantitative examination of ROIs for their diagnostic relevance, for example, using specialized quantitative ultrasound algorithms.

Over the past decade, artificial intelligence has significantly advanced automated medical image analysis [1], and deep learning (DL) models have substantially improved accuracy for image segmentation tasks. For the latter, a widely used DL architecture is the U-Net, which is built on convolutional neural networks and is characterized by its name-giving encoder-decoder shape [2]. The U-Net is capable of effectively capturing contextual features while preserving spatial information, even with relatively small datasets. It was originally developed for biomedical image segmentation by Ronneberger et al. in 2015 [3] and has since been adapted and refined by many research groups across different fields [2], [4].

In this study, we employ deep neural networks with a U-Net architecture for the segmentation of B-mode ultrasound images of the *vastus lateralis* muscle. A dataset of 130 images from 10 subjects with corresponding ground-truth segmentation masks was constructed. To address the challenges of low-data availability and image variability, we hypothesize that providing a deep learning model with an explicit, intensity-invariant feature map can serve as a powerful inductive bias. We engineer such a feature using monogenic phase asymmetry, which is designed to highlight anatomical structures like tissue boundaries, muscle fibers, and fascia, regardless of image contrast. We posit that this feature-augmented, lightweight model can outperform larger, more complex architectures that rely solely on representation learning from limited data. For benchmarking the performance of our models, the popular state-of-the-art U-Net implementation nnU-Net v2 from MIC-DKFZ was used [4].

Material and Methods

Data Acquisition and Preparation: Time-resolved ultrasound echo data (RF data) of the *vastus lateralis* muscle of ten healthy male volunteers were recorded with a 5 – 11 MHz linear array transducer from Telemed (LF11-5H60-A3). The field of view of the resulting B-Mode images is approximately 44x40 mm^2 (see Fig. 1 for the measurement setup and top left of Fig. 2 for an example B-Mode image). Three of the ten volunteers were measured twice, resulting in a total of 13 measurements. During each dynamic measurement, the subjects tensed their thigh muscles to 20 % of their predetermined maximum voluntary contraction (MVC), following a trapezoidal MVC curve (see Fig. 1). Each measurement lasted 40 s and the RF data was recorded at a frame rate of 100 Hz. The tensioning of the muscle visibly changes muscle boundaries and muscle fascicles configuration.

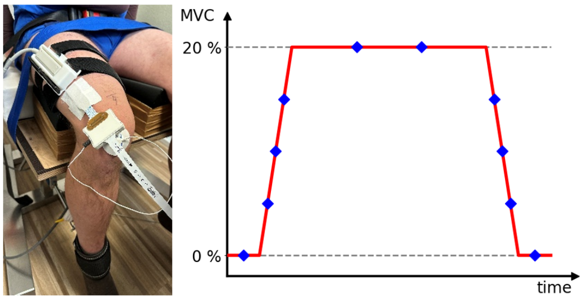

Fig. 1: Left: subject on a measurement seat with an ultrasound transducer strapped around the upper thigh and a force sensor around the ankle. Right: MVC curve which the subject follows during the measurement. Blue diamonds indicate the 10 points in time where the images were taken for the dataset.

For the training and test dataset, 10 frames were extracted from each measurement at points in time where the most variance between these frames is expected (see Fig. 1). In total, this resulted in 130 images, which were manually labeled with polygon-shaped binary segmentation masks (for examples, see Fig. 4). A set of 30 instances from 3 individual subjects was held out for final model evaluation.

U-Net Model Architecture: The deep neural networks were developed using Python 3.11, TensorFlow 2.18, and Keras 3.8. The U-Net architecture features a distinctive U-shaped encoder-decoder structure. The encoder extracts spatial features through convolutional blocks, while the decoder translates these into binary segmentation masks using 2D Up-Convolution and skip connections to preserve spatial information (see [3] for more details). As larger images did not significantly enhance model performance, the images were scaled down to 128x128 pixels to increase training speed. The most relevant hyperparameters were optimized through grid search, resulting in a network depth of 4, initial number of filters of 8, a dropout rate of 0.3 between network layers, and swish activation. In total, our model has approximately 390,000 trainable parameters, which is very few in the context of deep learning, making it lightweight and suitable for real-time applications. For training, a batch size of 1, the Dice loss, and the Adam optimizer with a learning rate of 0.001 were used.

Feature Engineering with Monogenic Phase Asymmetry: To provide the segmentation model with an explicit representation of structural information that is robust to intensity variations, we employed a feature engineering pipeline based on the monogenic signal. The monogenic signal, as introduced by Felsberg & Sommer [5], is a powerful mathematical framework for analyzing multidimensional signals such as images. It serves as a 2D generalization of the 1D analytic signal, a concept widely used in signal processing. The monogenic signal replaces the 1D Hilbert transform with the Riesz transform constructing a 3-component signal, combining the original image with the two components of its Riesz transform. This signal decomposes the image at each point into three fundamental, locally-defined properties: local amplitude (representing energy), local phase (representing structure), and local orientation (representing geometry) [6].

A key property of this decomposition is the *split of identity*, where local phase is invariant to changes in the signal's energy (i.e., image brightness and contrast), while the local amplitude is directly representative of that energy [5]. This characteristic makes phase-based features exceptionally well-suited for analyzing medical images like ultrasound, where absolute intensity values can be unreliable due to operator-dependent settings and physical artifacts, but the underlying anatomical structures remain consistent. The local phase effectively highlights line-like structures and edges, which correspond directly to the muscle fascicles and fascial planes that are critical for accurate segmentation. Notably, this approach is conceptually consistent with the formation of B-mode images themselves, which rely on the 1D analytic signal to demodulate the RF data and extract the signal envelope [6], [7].

From the local phase information, a more specialized feature known as phase asymmetry can be derived, which acts as a highly robust detector for edges and lines, particularly in low-contrast areas [8]. In the context of muscle ultrasound, this means it generates a map that sharply accentuates the muscle boundaries (aponeuroses) and the internal linear texture of the muscle fibers (fascicles). For this study, the phase asymmetry map was generated from the envelope of the original RF data for each corresponding B-mode image. This map, containing explicit and intensity-invariant structural information, was then used as the second input channel for the U-Net. This was implemented by setting the B-mode image and the corresponding phase asymmetry map to the red and green channels of an image, respectively, and feeding this red-green (RG) image into the U-Net, as shown in Fig. 2. In the following, we refer to this extension as the RG-UNet.

Benchmarking with nnU-Net and Evaluation Metrics: In order to have a benchmark for the performance of our models, we trained the popular self-configuring U-Net implementation nnU-Net v2 from the Division of Medical Image Computing of the German Cancer Research Center (MIC-DKFZ), which is publicly available on GitHub [4]. For the training with our dataset, the standard self-configuration was used.

DOI 10.5162/Ultrasonic2025/E4-a1

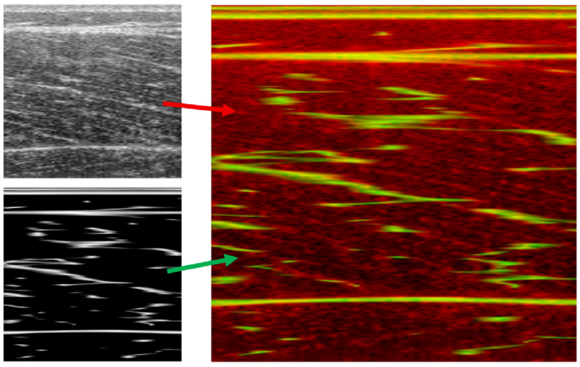

Fig. 2: The B-mode image (upper left) and its corresponding monogenic phase asymmetry map (lower left) are assigned to the red and green channels, respectively, of the resulting RG input image (right).

For the evaluation of segmentation algorithms, two standard metrics are most commonly used: Intersection over Union (IoU) and the Dice coefficient (DC). Both metrics measure the similarity between ground truth and predicted binary segmentation masks, ranging from 0 (no similarity) to 1 (identity).

Results

The U-Net, RG-UNet, and the nnU-Net were evaluated on the holdout test dataset of 30 images from 3 subjects. The resulting DC and IoU values along with their standard deviations are given in Tab. 1. The test scores are further illustrated in a boxplot in Fig. 3, and Fig. 4 provides a qualitative comparison of the predicted segmentation masks of all three models for 4 test instances.

Tab. 1: IoU and DC test scores with corresponding standard deviations for the three models.

Model	IoU	DC
nnU-Net	0.77 ± 0.15	0.86 ± 0.10
U-Net	0.89 ± 0.06	0.94 ± 0.04
RG-UNet	0.93 ± 0.06	0.96 ± 0.04

Both the U-Net and RG-UNet perform very well on the test dataset compared to the nnU-Net, where the RG-UNet exhibits the highest median and mean scores, and the U-Net has the most uniform distribution of test scores without any outliers. The nnU-Net is usually trained with much larger datasets and more often used in the context of MRI and CT imaging, so the comparison here is biased. In our configuration, it had around two orders of magnitude more trainable parameters than our U-Nets ($\sim$45,000,000 vs. $\sim$390,000), making it more prone to overfitting, especially with a small training dataset. More fundamentally, its performance decline highlights the challenge of pure representation learning when data is scarce. Without a sufficient number of examples, the model struggles to learn the necessary invariances to contrast and texture.

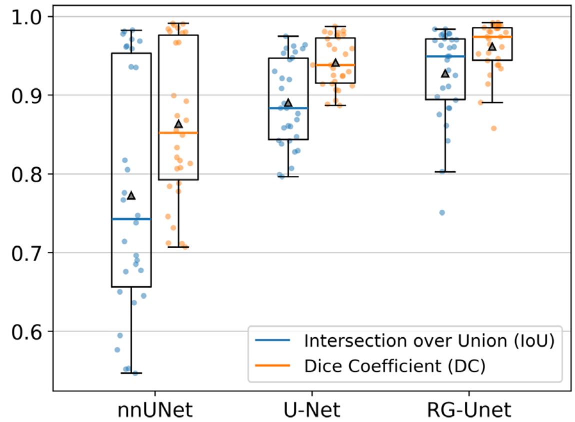

Fig. 3: Boxplot comparing the IoU and DC scores of the three evaluated models for the 30 test instances. Median values are shown as colored lines, mean values as black triangles.

In contrast, our RG-UNet is provided with a powerful, domain-relevant inductive bias through the phase asymmetry channel. This pre-engineered feature explicitly encodes the structural information, guiding the lightweight network to a more robust and generalizable solution in a data-efficient manner. From Fig. 3 and Fig. 4, it can be seen that for test images from one subject, the nnU-Net performed as well as and even better than the other two models, but it could not generalize nearly as well to the images of the other two subjects.

Overall, for test instances with clear muscle-skin separation, the U-Net and RG-UNet models consistently achieved IoU scores well above 0.9 (third row in Fig. 4). However, in cases with unclear separations, low signal-to-noise ratio, or bright echoes in the skin fat layer, some predicted masks showed significant artifacts such as holes and bulges (first, second, and fourth row in Fig. 4). In most of these challenging cases, the RG-UNet excelled over the standard U-Net.

Conclusion and Outlook

The U-Net demonstrated itself to be a lightweight, efficient, and highly capable deep learning model for segmenting muscle ultrasound B-mode images, even with a relatively small dataset. Incorporating monogenic phase asymmetry maps into the U-Net architecture as an additional input channel further improved

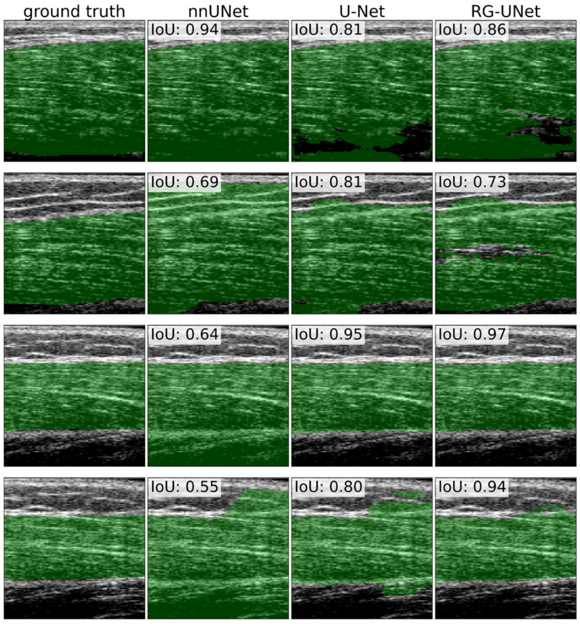

Fig. 4: Qualitative comparison of the predicted segmentation masks (green overlays) produced by the three models for 4 test instances.

its performance, underscoring the value of integrating engineered features that provide a strong inductive bias.

A larger and more diverse dataset could unlock the full potential of the nnU-Net, which will be explored in future studies. Additional measurements involving more subjects are already planned. Furthermore, individual frames from a single dynamic ultrasound measurement show a high degree of temporal correlation. Modifying the U-Net architecture to process time-resolved images could further enhance its generalizability. Here, the nnU-Net may perform significantly better, as it can be natively fed with 3D input data and is also able to handle sparse and so-called scribble annotations [9].

Looking forward, while this study demonstrates the profound impact of a fixed, engineered feature, a promising avenue for future research lies in learning these features end-to-end. Recent advancements in trainable monogenic layers allow for the optimization of filter parameters, such as scale and bandwidth, directly within the deep learning framework [10]. Integrating such a trainable layer could further enhance model performance and generalization by automatically adapting the feature extraction process to the specific characteristics of the dataset, representing a logical next step that builds upon the foundational findings presented here.

References

[1] L. Pinto-Coelho, 'How artificial intelligence is shaping medical imaging technology: A survey of innovations and applications,' *Bioengineering*, vol. 10, no. 12, p. 1435, 18th Dec. 2023. DOI: 10.3390/bioengineering10121435.

[2] R. Azad *et al.*, 'Medical image segmentation review: The success of u-net,' *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 10 076–10 095, Dec. 2024. DOI: 10.1109/TPAMI.2024.3435571.

[3] O. Ronneberger, P. Fischer and T. Brox, 'U-net: Convolutional networks for biomedical image segmentation,' in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells and A. F. Frangi, Eds., vol. 9351, Series Title: Lecture Notes in Computer Science, Cham: Springer International Publishing, 2015, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28.

[4] F. Isensee, P. F. Jaeger, S. A. A. Kohl, J. Petersen and K. H. Maier-Hein, 'nnU-net: A self-configuring method for deep learning-based biomedical image segmentation,' *Nature Methods*, vol. 18, no. 2, pp. 203–211, Feb. 2021. DOI: 10.1038/s41592-020-01008-z.

[5] M. Felsberg and G. Sommer, 'The monogenic signal,' *IEEE transactions on signal processing*, vol. 49, no. 12, pp. 3136–3144, 2002.

[6] C. Wachinger, T. Klein and N. Navab, 'The 2d analytic signal for envelope detection and feature extraction on ultrasound images,' *Medical image analysis*, vol. 16, no. 6, pp. 1073–1084, 2012.

[7] C. P. Bridge, 'Introduction to the monogenic signal,' *arXiv preprint arXiv:1703.09199*, 2017.

[8] K. Mei, B. Hu, B. Fei and B. Qin, 'Phase asymmetry ultrasound despeckling with fractional anisotropic diffusion and total variation,' *IEEE Transactions on Image Processing*, vol. 29, pp. 2845–2859, 2019.

[9] K. Gotkowski *et al.*, *Embarrassingly simple scribble supervision for 3d medical segmentation*, 2024. arXiv: 2403.12834.

[10] A. Kimbowa, A. Parmar, M. Badii, D. Liu, M. Harkey and I. Hacihaliloglu, 'Mono2d: A trainable monogenic layer for robust knee cartilage segmentation on out-of-distribution 2d ultrasound data,' *arXiv preprint arXiv:2503.09050*, 2025.