Article

Adaptive Evolutionary Optimization of Deep Learning Architectures for Focused Liver Ultrasound Image Segmentation

Ali Zifan [1,]*[🔵], Katelyn Zhao [1], Madilyn Lee [1], Zihan Peng [1], Laura J. Roney [1], Sarayu Pai [1], Jake T. Weeks [2], Michael S. Middleton [2], Ahmed El Kaffas [2], Jeffrey B. Schwimmer [3,4] and Claude B. Sirlin [2]

[1] Division of Gastroenterology and Hepatology, University of California San Diego, San Diego, CA 92093, USA; k5zhao@ucsd.edu (K.Z.); mal069@ucsd.edu (M.L.); z7peng@ucsd.edu (Z.P.); ljroney@ucsd.edu (L.J.R.); s2pai@ucsd.edu (S.P.)

[2] Liver Imaging Group, Department of Radiology, University of California San Diego, San Diego, CA 92093, USA; j1weeks@health.ucsd.edu (J.T.W.); msm.ucsd.edu@gmail.com (M.S.M.); aelkaffas@health.ucsd.edu (A.E.K.); csirlin@health.ucsd.edu (C.B.S.)

[3] Department of Pediatrics, Division of Gastroenterology, Hepatology, and Nutrition, University of California San Diego School of Medicine, La Jolla, CA 92093, USA; jschwimmer@health.ucsd.edu

[4] Department of Gastroenterology, Rady Children's Hospital San Diego, San Diego, CA 92123, USA

[*] Correspondence: azifan@health.ucsd.edu

Abstract: Background: Liver ultrasound segmentation is challenging due to low image quality and variability. While deep learning (DL) models have been widely applied for medical segmentation, generic pre-configured models may not meet the specific requirements for targeted areas in liver ultrasound. Quantitative ultrasound (QUS) is emerging as a promising tool for liver fat measurement; however, accurately segmenting regions of interest within liver ultrasound images remains a challenge. **Methods:** We introduce a generalizable framework using an adaptive evolutionary genetic algorithm to optimize deep learning models, specifically U-Net, for focused liver segmentation. The algorithm simultaneously adjusts the depth (number of layers) and width (neurons per layer) of the network, dropout, and skip connections. Various architecture configurations are evaluated based on segmentation performance to find the optimal model for liver ultrasound images. **Results:** The model with a depth of 4 and filter sizes of [16, 64, 128, 256] achieved the highest mean adjusted Dice score of 0.921, outperforming the other configurations, using three-fold cross-validation with early stoppage. **Conclusions:** Adaptive evolutionary optimization enhances the deep learning architecture for liver ultrasound segmentation. Future work may extend this optimization to other imaging modalities and deep learning architectures.

Keywords: ultrasound liver segmentation; deep learning optimization; evolutionary genetic algorithm

1. Introduction

Metabolic dysfunction-associated steatotic liver disease (MASLD) affects approximately 10% of children and, if left untreated, may progress to its more advanced form, metabolic dysfunction-associated steatohepatitis (MASH), and to long-term complications such as diabetes, cardiovascular disease, cirrhosis, and liver cancer [1–5]. Magnetic resonance imaging proton-density fat fraction (MRI-PDFF) is widely recognized as the non-invasive reference standard to assess hepatic steatosis, the hallmark feature of MASLD [6,7]; however, its availability is limited due to its infrastructure requirements, and costs [8], especially in resource-limited regions worldwide. In contrast, ultrasound (US) is safe,

non-invasive, more affordable, and more widely available, but may underdiagnose mild steatosis due to operator and machine dependence, and reliance on reader assessment [9,10].

Appreciation of the limitations of US has sparked growing interest in developing quantitative ultrasound (QUS) as an alternative technology to assess possible MASLD [11,12]. Although QUS has shown promise to be more accurate and precise than US, additional research is needed to fully validate QUS [13]. In one approach, the liver boundaries are determined using manual US liver segmentation, which is time-consuming and labor-intensive, and then, hepatic steatosis and perhaps other histologic features of MASLD are assessed using the QUS methodology. An important step to improve QUS liver segmentation is to automate liver boundary delineation, as this is currently performed manually by radiologists.

Deep learning techniques have proven efficacious in segmenting liver MRI and CT [14–23], but not yet US images [1,24–32]. Moreover, existing liver US approaches are limited and have seen only limited mainstream application due to the lack of sufficient data or reproducibility [33]. The aim of this study is to address these challenges by advancing toward a fully automated targeted-field-of-interest liver US segmentation by using genetic evolutionary algorithms to determine the optimal neural network architecture.

2. Materials and Methods

To achieve automated and robust targeted-region-of-interest ultrasound (US) liver segmentation, we introduce a new method to optimize an established encoder–decoder model such as the U-Net [34] deep learning (DL) architecture using evolutionary algorithms (EAs), though it should be noted that any other model could be used, as the goal is architectural optimization. Evolutionary algorithms are optimization techniques inspired by natural selection processes that iteratively refine solutions by simulating evolution. EAs provide a powerful framework for exploring and refining network architecture; they naturally lend themselves to parallel processing and help speed up the search process. Genetic algorithms (GAs), a subset of EAs, use mechanisms like selection, crossover, and mutation to evolve solutions over generations. GAs are particularly useful when computation times for individual evaluations are long, making them well suited to solving problems that require global search, robustness, and flexibility [35]. Natural selection is a process in biology where organisms better adapted to their environment tend to survive and reproduce, passing on advantageous traits to future generations. By simulating natural selection, evolutionary algorithms (EAs) iteratively evolve deep learning (DL) architectures to enhance liver segmentation performance. Mutation is a process where random changes are introduced to the architecture, allowing new variations that may improve performance. Crossover combines features from two parent architectures to create offspring that could potentially inherit strengths from both. Migration involves transferring solutions or information between populations, encouraging diversity in the search space. Selection is the process of choosing the best-performing architectures to pass their features to the next generation. Through these mechanisms, candidate architectures are generated and evaluated based on segmentation accuracy, allowing U-Net models to adapt and optimize for liver ultrasound segmentation without manual tuning.

2.1. Experimental Setup

Ultrasound data were acquired from 30 children (mean age: 13 ± 2.4 years). Among the participants, 12 (40%) had a clinical diagnosis of MASLD, 17 (56.7%) were at risk for MASLD, and 1 (3.3%) had neither. We used the C1-6 curved array probe of a GE Logiq E10 system (GE HealthCare, Chicago, IL, USA). The study was approved by the UCSD IRB. Parents gave written consent and participants gave written assent. Transverse B-

mode images of the right lobe of the liver were acquired through an intercostal window in both the fundamental and harmonic modes by one of two study-trained registered diagnostic medical sonographers. For each child, the sonographer selected the settings that optimized liver visualization in their judgment. These settings included gain, time-gain compensation, depth, and transmit center frequency (3.0 MHz or 4.0 MHz for fundamental mode, 3.0 MHz or 4.5 MHz for harmonic mode). Transverse scanning was chosen to standardize the imaging protocol and ensure consistent visualization of liver boundaries, which are critical for evaluating segmentation performance. While we acknowledge its limitations in visualizing deeper liver segments (e.g., segments 6 and 7), this approach minimizes variability introduced by operator-dependent factors, such as probe angle and patient positioning. Future work will explore integrating oblique or intercostal scanning views to address challenges related to sound refraction and ultrasound attenuation through the abdominal rectus muscle.

Liver boundaries on each B-mode image were segmented manually by a trained image analyst under the supervision of a radiologist using the ITK-SNAP segmentation tool. The field of interest was drawn to capture as much liver parenchyma as possible while avoiding liver edges, shadows, dropout, and other artifacts. No effort was made to avoid blood vessels in the field of interest, as also our previous preliminary data indicated that vessel removal did not affect the results.

The B-mode images were subsequently loaded on a Dell Precision T7910, Dual Intel Xeon Processor E5-2687W v4, NVIDIA Quadro M6000 24 GB, 256 GB RAM. OpenAI's generative AI tools were used to assist with language editing and grammar correction. The tools were employed exclusively for refining the text and did not contribute to the conceptualization, data analysis, or interpretation of the study results. We used Python 3.8 (Python Software Foundation, https://www.python.org), and Keras [36] for our EA implementation, which also facilitated multiprocessing, enabling the simultaneous training of multiple genomes for the liver segmentation task.

2.2. Evolutionary Genomic Optimization

We employed a multi-population evolutionary approach to enhance the training of our predictive U-Net model (see Figure 1). This method involved simultaneously training across multiple genomic subpopulations, allowing us to leverage their unique characteristics and improve the model's robustness and generalizability.
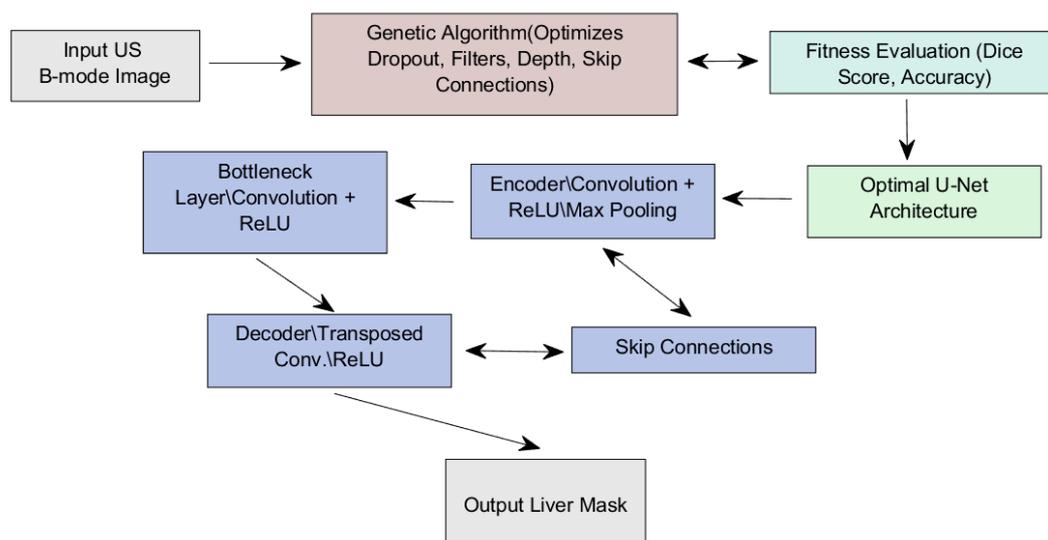

Figure 1. A flow diagram of the proposed genomic optimization of the U-Net model.

We used parallel genomic training sessions with varying depths (see Table 1) to capture a broader spectrum of features and interactions. By systematically varying the model's depth, dropout rate, and skip connections (set to True or False), we were able to explore different levels of complexity in the abdominal ultrasound images, ensuring that both shallow and deep representations of the genomic information were effectively learned. Moreover, we integrated migration techniques to facilitate the transfer of knowledge between subpopulations. This enabled the model to retain learned representations from shallower depths and apply them to deeper models, enhancing its ability to recognize and classify patterns across diverse genomic landscapes. The novelty of this approach is the combined effect of these methodologies, which allows us to create a flexible and adaptable deep learning model architecture that could effectively interpret the complex relationships within the liver US genomic data.

- **Genome Representation**: We represent each genome as a dictionary containing the following hyperparameters:
  - **Dropout Rate**: p_d, where $p_d \in [0, 0.5]$.
  - **Filter Sizes**: a list of integers representing the number of filters in each layer (see Table 1), $F = [f_1, f_2, f_3, f_4]$.
  - **Depth**: d, representing the number of layers, where $d \in [2, 5]$.
  - **Use Skip Connections**: a Boolean flag u_s, indicating whether skip connections are included.

- **Fitness Function:** the fitness of each U-Net genome is evaluated based on the Dice coefficient:

Table 1. Filter sizes explored during the evolutionary optimization of the U-Net architecture.

Depth
3
[8, 16, 128], [8, 32, 128], [16, 32, 128]
[8, 64, 256], [64, 128, 256], [16, 32, 64], [32, 64, 128]
4
[16, 32, 64, 128], [64, 128, 256, 512], [8, 32, 128, 256]
[8, 64, 128, 512], [16, 64, 128, 256], [32, 64, 128, 256]
5
[16, 32, 64, 128, 256], [32, 64, 128, 256, 512], [8, 32, 128, 256, 512],
[8, 64, 256, 512, 1024], [64, 128, 256, 512, 1024]
6
[32, 64, 128, 256, 512, 1024], [8, 16, 64, 128, 256, 512], [16, 32, 64, 128, 256, 512]
[8, 32, 128, 256, 512, 1024], [16, 32, 128, 256, 512, 1024]

$$Dice\left(y_{true}, y_{pred}\right) = \frac{2 \cdot \left|y_{true} \cap y_{pred}\right|}{\left|y_{true}\right| + \left|y_{pred}\right|} + \epsilon \tag{1}$$

where ϵ is a small constant to prevent division by zero. Additionally, we define the ***average fitness*** across the population:

$$Avg\ Fitness = 1/N \sum_{i=1}^{N} Fitness_i \tag{2}$$

where N is the population size.

- **Selection:** The population is sorted based on fitness scores, and the top half of the genomes is retained for the next generation. The selected genome can be represented as

$$Winner\ Genome = argmax(Fitness_i), \quad i \in [1, N] \tag{3}$$

- **Crossover:** Two parent genomes are randomly selected to produce offspring through the following rules:

  - The dropout rate and depth are averaged:

$$p_{d,child} = \frac{p_{d,parent1} + p_{d,parent2}}{2}$$
$$d_{child} = round\left(\frac{p_{d,parent1} + p_{d,parent2}}{2}\right) \tag{4}$$

  - The filter sizes are averaged and rounded to the nearest integer:

$$f_{i,child} = round\left(\frac{f_{i,parent1} + f_{i,parent2}}{2}\right), i \in [1, N] \tag{5}$$

  - The skip connection flag is randomly selected from the parents.

- **Migration:** Facilitates the transfer of knowledge between subpopulations:

  - After every few generations, a certain percentage of genomes are migrated between subpopulations. This can be represented as

$$p_{d,migrated} = (1 - migration\ rate) \cdot p_{d,original} + migration\ rate \cdot p_{d,source} \tag{6}$$

  - This influences the fitness evaluation and crossover processes.

- **Mutation**: Random mutations are applied to introduce variability:

  - With a probability of 10%, the dropout rate is perturbed:

$$p_{d,mutated} = p_{d,child} + \Delta p, \quad \Delta p \sim u(-0.05, 0.05) \tag{7}$$

  - With a probability of 10%, each filter size is adjusted by ± 8 filters, ensuring the values stay within the valid range:

$$f_{i,mutated} = clip\left(f_{i,child} + \Delta f, min_f, max_f\right), \quad \Delta f \sim u(-8, 8) \tag{8}$$

In our setting, we select the top-performing genome from each subpopulation (depth) based on fitness (Dice) scores, transferring its dropout rate and skip connection configuration to other subpopulations. Next, when a genome migrates, new filter sizes are generated to align with the target depth's number of layers, ensuring compatibility. Subsequently, each depth retains only the top 3 genomes after migration, maintaining a focused search within each subpopulation while allowing beneficial traits to spread across different depths.

- **Boundary Constraints:** To maintain valid parameter ranges, we apply clipping for the filter sizes and dropout rates:

$$f_{i,constrained} = clip(f_i, f_{min}, f_{max}) p_{d,constrained} = clip(p_d, 0, 0.5) \tag{9}$$

- **Depth Penalization:** We also penalize deeper networks to prevent overfitting and manage the trade-off between model complexity and performance (avoiding extra training parameters):

$$Adjusted\ Dice\ score = Best\ Dice \cdot \frac{1}{d \cdot p}, \tag{10}$$

where *d: depth* and *p: penalty factor.*

- **Convergence Criteria:** We run the algorithm for a predefined number of generations (e.g., 30 epochs), *gen*, or until the change in average fitness across generations is below a threshold ϵ_c:

$$Convergence = if \left| Avg\ Fitness_{gen} - Avg\ Fitess_{gen-1} \right| < \epsilon_c \qquad (11)$$

- **Training Process**

The best genome identified by the GA is used to configure the U-Net model. The model is trained on the preprocessed dataset with the following loss function:

$$Loss = -\left(y_{true}\log\left(y_{pred} \right) + (1 - y_{true})\log\left(1 - y_{pred} \right) \right) \qquad (12)$$

Next, the model is optimized using an Adam optimizer with a learning rate of α set at 1×10^{-4}. The performance of the model is evaluated using the Dice coefficient as the primary metric.

3. Results

In our optimization approach, the first step involved genome optimization to identify the best architecture, as described previously. This involved fine-tuning key parameters, including filter sizes, depth, skip connections, and dropout rates. To expedite the identification of an initial winner, we employed a smaller number of epochs (30) during this phase, utilizing an 80-20 train–test split for model evaluation on 627 analyst-labeled 256-by-256 liver ultrasound images (see Figure 2). The optimization was facilitated through a combination of crossover, mutation, and migration techniques, along with penalizing larger depths using a factor of 0.1, thereby encouraging the exploration of more efficient architectures. The second step utilized the optimized architecture identified in the first phase, increasing the number of sampling epochs to 300 (with early stoppage) to achieve further refinement of the winning architecture. By passing the best-performing model from one generation to the next (10 generations), we introduced additional modifications through mutation and crossover, which led to the evolution of increasingly effective models. The proposed segmentation framework prioritizes time and cost efficiency: the evolutionary optimization process converged within 48 GPU hours, while the optimized model processes each image in under 0.1 s on a standard GPU.

The phase 1 (i.e., multi-population genomic optimization) results revealed that the model with a depth of 4 and filter sizes of [16, 64, 128, 256] emerged as the top performer, achieving an adjusted Dice score of 0.859 on 627 images with a size of 256 by 256. This score not only outperformed the other configurations tested but also ranked as the best architecture among all 10 generations (see Table 2). In comparison, the depth 3 model with filter sizes of [16, 32, 128] achieved an adjusted Dice score of 0.685. The depth 5 model, configured with filter sizes of [16, 32, 64, 128, 256], scored 0.828, while the depth 6 model, utilizing filter sizes of [32, 64, 128, 256, 512, 1024], yielded an adjusted score of 0.826. An interesting collateral of these results is that increasing depth and complexity does not inherently lead to improved performance, as indicated by the lower scores of the depth 5 and depth 6 models compared to the depth 4 configuration. Naturally, as generations progressed, the architectures exhibited noticeable improvements in performance, underscoring the effectiveness of the proposed optimization strategy. The proposed evolutionary pipeline successfully identified the depth 4 architecture with filter sizes of [16, 64, 128, 256] and a dropout rate of 0.1464 (with skip connections) as the optimal choice for the segmentation of the US liver dataset. This configuration not only demonstrated superior performance but also consistently ranked among the top architectures across all

generations. Finally, the optimal genome was used to train the U-Net model with early stopping (max 300 epochs) to obtain the final model. This resulted in a mean Dice score of 0.92 with a standard deviation of 0.00124 across three folds. See Table 2 and Figure 3 for the sample results of the prediction model applied on six random US samples in the dataset. Representative examples of segmented liver boundaries were overlaid on the original ultrasound images to illustrate the model's performance. These visualizations highlight the method's ability to accurately delineate liver parenchyma despite challenges in low-quality ultrasound data.
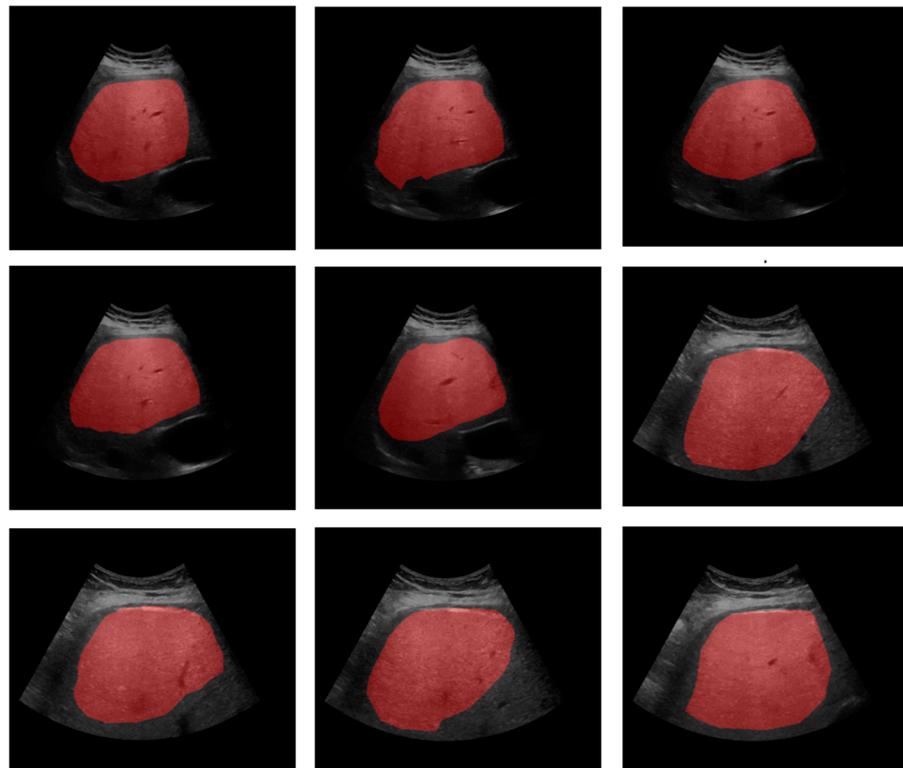

Figure 2. Nine samples of different subjects with superimposed ground truth segmentation.

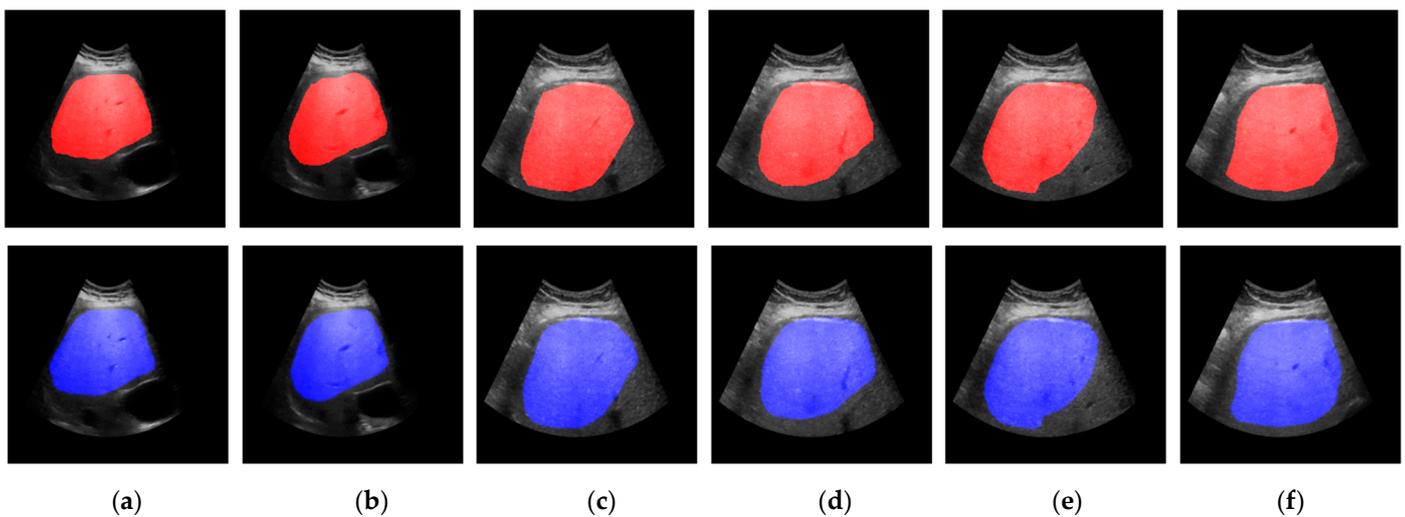| **(a)** | **(b)** | **(c)** | **(d)** | **(e)** | **(f)** |

Figure 3. The segmentation results showing the original images with ground truth segmentation overlaid in red at the top and predicted segmentation in blue at the bottom for three samples: (**a**–**f**).

Table 2. U-Net architectures [1] at varying depths across ten generations of evolutionary optimization.

Generation	Depth and Filter Sizes/Adjusted Scores
Gen 1	3: [8, 64, 256] (0.66874), 4: [64, 128, 256, 512] (0.85492), 5: [16, 32, 64, 128, 256] (0.86036), 6: [32, 64, 128, 256, 512, 1024] (0.85848)
Gen 2	3: [64, 128, 256] (0.49106), 4: [64, 128, 256, 512] (0.79684), 5: [16, 32, 64, 128, 256] (0.79872), 6: [32, 64, 128, 256, 512, 1024] (0.87574)
Gen 3	3: [8, 64, 256] (0.75624), 4: [64, 128, 256, 512] (0.86317), 5: [16, 32, 64, 128, 256] (0.84136), 6: [32, 64, 128, 256, 512, 1024] (0.77848)
Gen 4	3: [8, 64, 256] (0.60629), 4: [64, 128, 256, 512] (0.88488), 5: [16, 32, 64, 128, 256] (0.77838), 6: [32, 64, 128, 256, 512, 1024] (0.79681)
Gen 5	3: [64, 128, 256] (0.67891), 4: [64, 128, 256, 512] (0.85566), 5: [16, 32, 64, 128, 256] (0.79540), 6: [32, 64, 128, 256, 512, 1024] (0.78476)
Gen 6	3: [8, 64, 256] (0.85739), 4: [16, 64, 128, 256] (0.76498), 5: [16, 32, 64, 128, 256] (0.83489), 6: [32, 64, 128, 256, 512, 1024] (0.83680)
Gen 7	3: [8, 64, 256] (0.63057), 4: [16, 64, 128, 256] (0.83652), 5: [16, 32, 64, 128, 256] (0.83393), 6: [8, 32, 128, 256, 512, 1024] (0.76690)
Gen 8	3: [8, 64, 256] (0.76110), 4: [16, 64, 128, 256] (0.80789), 5: [16, 32, 64, 128, 256] (0.73106), 6: [32, 64, 128, 256, 512, 1024] (0.84866)
Gen 9	3: [16, 32, 128] (0.43480), 4: [16, 64, 128, 256] (0.81051), 5: [64, 128, 256, 512, 1024] (0.81905), 6: [8, 32, 128, 256, 512, 1024] (0.78400)
Gen 10	3: [16, 32, 128] (0.68542), 4: [16, 64, 128, 256] (0.85939), 5: [16, 32, 64, 128, 256] (0.82776), 6: [32, 64, 128, 256, 512, 1024] (0.82564)

[1] The results from the optimization process in the study on U-Net architecture for US liver segmentation. The adjusted scores represent the performance metric for each model configuration, with higher scores indicating better segmentation accuracy. The filter sizes are specified in brackets, and the adjusted Dice scores in parenthesis.

4. Discussion

We employed evolutionary algorithms to optimize the architecture of a deep learning model, specifically a U-Net, for segmenting the liver boundary in abdominal ultrasound images from children. The proposed segmentation method offers several critical contributions to clinical applications: (A) Liver Fat Quantification: By enabling accurate delineation of liver parenchyma, this method ensures precise measurements of QUS parameters such as the attenuation coefficient and backscatter coefficient. These parameters are foundational for diagnosing conditions like MASLD or liver fibrosis. (B) Early Diagnosis of Liver Cirrhosis: The proposed segmentation allows for the automated extraction of liver texture and morphological features, which serve as biomarkers for cirrhosis detection. (C) Enhanced Workflow Consistency: By automating ROI selection, our method reduces operator dependency, improving the reproducibility and reliability of QUS analysis.

Moreover, by iteratively refining parameters such as dropout rates, filter sizes, depth, and skip connections, we demonstrated that this approach could lead to optimized architectures. While our study focused on the U-Net architecture, the proposed optimization methodology can be readily applied to any network architecture which has an encoder–decoder design, offering a flexible and effective solution for improving segmentation outcomes across various domains. Our results indicate that evolutionary optimization has the potential to substantially boost the accuracy and robustness of deep learning models in ultrasound liver segmentation. Moreover, by leveraging open-source tools and eliminating the need for proprietary software, implementation costs were minimized. Furthermore, the automated segmentation reduces reliance on manual annotations, saves valuable time for clinicians, and ensures the framework's scalability and feasibility for diverse clinical and operational settings.

The main goal of our proposed method was to optimize deep learning architectures to maximize segmentation quality and accuracy, particularly for challenging low-quality ultrasound images. While the current architecture achieves competitive processing speed,

further optimizations are possible. For example, one could adopt lightweight design strategies such as depthwise separable convolutions, which split standard convolutions into depthwise and pointwise operations. This could reduce the computational burden while retaining performance. Or we could also explore techniques like group convolutions, channel shuffling, or even attention mechanisms to further improve efficiency by reducing redundancy, with an augmented focus on computation. While such adaptations are beyond the scope of this study, they represent directions for enhancing speed without compromising segmentation quality, particularly in real-time or resource-constrained clinical settings.

However, our study is not without limitations. Other limitations include using a single transverse image, focusing on children only, employing one transducer (GE LOGIQ E10) only, and utilizing an intercostal view only. One notable constraint is restricting the filter sizes to powers of two (as is commonly used, as they align better with computer architecture for optimal memory management and processing efficiency). Although this choice enhances computational performance and simplifies batch processing, leading to a more structured and scalable network design, it inherently constrains the search space and the random evolutionary nature of the method, potentially overlooking filter sizes that are not powers of two. In a pure evolutionary setting, filter sizes can vary widely, and architectural configurations need not necessarily adhere to a funnel shape or a strictly monotonic increase or decrease. However, filter sizes were restricted to powers of two to optimize computational efficiency and compatibility with GPU memory alignment. Additional experiments with unrestricted filter sizes demonstrated negligible performance improvements (<1% Dice score increase) while significantly increasing computational costs. This practical design choice ensures an optimal balance between performance and efficiency. Therefore, our future studies could benefit from a more expansive exploration of filter sizes, allowing for a more diverse set of architectures that may yield superior performance.

Furthermore, the mutation strategy employed in our genetic algorithm could also be further enhanced. While the primary focus of our genomic optimization was on optimizing filter size and depth, dropout and skip connections, alternative mutation strategies might lead to even better outcomes. However, exploring different mutation mechanisms was outside the scope of our paper, but remains an intriguing avenue for future research. Another drawback of using genetic algorithms is the considerable training time they entail, particularly given the computational complexity involved in evaluating multiple generations of architectures. To address this challenge, we leveraged the multiprocessing capabilities provided by Python's multiprocessing module, specifically using the 'Pool', 'Manager', and 'Lock' classes, which allowed us to simultaneously train different genomes, significantly reducing the overall computational burden and expediting the optimization process.

Finally, one could explore hybrid strategies. For example, while we applied genetic optimization for optimizing all genome features at once, an alternative strategy could be to integrate differential evolution [37] techniques for optimizing hyperparameters such as dropout rates and learning rates while confining genetic optimization specifically for only filter and depth optimization. This hybrid approach could offer a more comprehensive optimization framework, allowing for finer control over various aspects of the network architecture and training process, also including the learning rate itself as an additional search parameter, allowing us to achieve a balance between faster convergence and model stability, enhancing the overall training efficiency. Reproducibility and generalizability remain critical challenges in medical imaging studies. Factors such as vendor variability, acquisition protocols, and patient demographics influence reproducibility. Our future work will address these challenges by incorporating data from multiple ultrasound vendors, standardizing acquisition protocols, and including diverse patient populations with liver

pathologies. Importantly, the proposed methodology is not tied to specific devices, making it adaptable to various clinical and research settings.

5. Conclusions

In summary, our results suggest that genomic optimization using evolutionary algorithms is a highly promising avenue for optimizing deep learning architectures in field-of-interest liver ultrasound segmentation. This process allows the algorithmic models to adapt to the specific needs of the segmentations being performed. By utilizing parallelization, multiple architectural representations can be trained simultaneously on any encoder–decoder architecture, deriving the best-performing hyperparameters for a specific segmentation task. Though we applied our methods to liver ultrasound segmentation, this approach is generalizable and can be readily applied to the segmentation of other organs across different imaging modalities.

Author Contributions: Conceptualization, A.Z., C.B.S. and J.B.S.; methodology, A.Z.; software, A.Z.; validation, A.Z., K.Z. and Z.P.; formal analysis, A.Z., K.Z., Z.P. and J.T.W.; investigation, A.Z., K.Z., Z.P., L.J.R., M.L., M.S.M., A.E.K., S.P., J.T.W., J.B.S. and C.B.S.; resources, C.B.S., J.B.S. and A.Z.; data curation, J.T.W., C.B.S. and J.B.S.; writing—original draft preparation, A.Z.; writing—review and editing, K.Z., Z.P., L.J.R., J.T.W., M.L., A.E.K., S.P., J.B.S., M.S.M. and C.B.S.; visualization, A.Z.; supervision, A.Z.; project administration, C.B.S. and J.B.S.; funding acquisition, C.B.S. and J.B.S. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: The human investigation committee of the University of California, San Diego, approved the study protocol (IRB # 800790).

Informed Consent Statement: In the conducted study, the principle of informed consent was diligently adhered to, and consent was obtained from all participating subjects. Prior to their involvement, everyone was provided with comprehensive information about the study's purpose, procedures, potential risks, and benefits. They were given ample time to ask questions and clarify any uncertainties before voluntarily agreeing to participate. The informed consent process aimed to ensure that the subjects fully understood the implications of their involvement and that they willingly and freely consented to be part of this research.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to IRB.

Conflicts of Interest: Claude B. Sirlin reports payment to the institution for research grants from ACR, Bayer, Foundation of NIH, GE, Gilead, Pfizer, Philips, Siemens, and V Foundation; payment to the institution for lab service agreements with OrsoBio, Enanta, Gilead, ICON, Intercept, Nusirt, Shire, Synageva, and Takeda; payment to the institution for institutional consulting for BMS, Exact Sciences, IBM-Watson, and Pfizer; personal consulting for Altimmune, Ascelia Pharma, Blade, Boehringer, Epigenomics, Guerbet, Livivos, and Novo Nordisk; payment to self for royalties and/or honoraria from Medscape, Wolters Kluwer, and HealthProMatch; ownership of stock options in Livivos; an unpaid advisory board position at Quantix Bio; an executive position at Livivos (Chief Medical Officer, unsalaried position with stock options and stock) throughout 28 June 2023; a Principal Scientific Advisor position at Livivos (unsalaried position with stock options and stock) since 28 June 2023; payment to self for serving as a speaker for HealthProMatch; support for attending meetings and/or travel from Fundacion Santa Fe, Congreso Argentino de Diagnóstico por Imágenes, Stanford, Jornada Paulista de Radiologia, Ascelia Pharma, RSNA, Sociedad Radiológica de Puerto Rico, Hospital Español Auxilio Mutuo de Puerto Rico, Paris MASH, and the Liver Forum; membership (no payment) of the Data Safety Monitoring board for National Cancer Institute funded Early Detection Research Network; equipment loans to the institution from Butterfly, GE, Siemens, and Mayo; and the provision of contrast material to the institution from Bayer. Michael S. Middleton reports providing consultation to Alimentiv, Arrowhead, AutonomUS, Glympse, Immunobrain, Kowa, Livivos, Median, Novo

References

1. Ibrahim, M.N.; Blázquez-García, R.; Lightstone, A.; Meng, F.; Bhat, M.; El Kaffas, A.; Ukwatta, E. Automated fatty liver disease detection in point-of-care ultrasound B-mode images. *J. Med. Imaging (Bellingham)* **2023**, *10*, 034505. [CrossRef] [PubMed]

2. Engin, A. Nonalcoholic Fatty Liver Disease and Staging of Hepatic Fibrosis. *Adv. Exp. Med. Biol.* **2024**, *1460*, 539–574. [CrossRef]

3. Pouwels, S.; Sakran, N.; Graham, Y.; Leal, A.; Pintar, T.; Yang, W.; Kassir, R.; Singhal, R.; Mahawar, K.; Ramnarain, D. Non-alcoholic fatty liver disease (NAFLD): A review of pathophysiology, clinical management and effects of weight loss. *BMC Endocr. Disord.* **2022**, *22*, 63. [CrossRef]

4. Powell, E.E.; Wong, V.W.-S.; Rinella, M. Non-alcoholic fatty liver disease. *Lancet* **2021**, *397*, 2212–2224. [CrossRef] [PubMed]

5. Schwimmer, J.B.; Deutsch, R.; Kahen, T.; Lavine, J.E.; Stanley, C.; Behling, C. Prevalence of fatty liver in children and adolescents. *Pediatrics* **2006**, *118*, 1388–1393. [CrossRef] [PubMed]

6. Starekova, J.; Hernando, D.; Pickhardt, P.J.; Reeder, S.B. Quantification of Liver Fat Content with CT and MRI: State of the Art. *Radiology* **2021**, *301*, 250–262. [CrossRef] [PubMed]

7. Raptis, D.A.; Fischer, M.A.; Graf, R.; Nanz, D.; Weber, A.; Moritz, W.; Tian, Y.; Oberkofler, C.E.; Clavien, P.A. MRI: The new reference standard in quantifying hepatic steatosis? *Gut* **2012**, *61*, 117–127. [CrossRef]

8. Bresnahan, R.; Duarte, R.; Mahon, J.; Beale, S.; Chaplin, M.; Bhattacharyya, D.; Houten, R.; Edwards, K.; Nevitt, S.; Maden, M.; et al. Diagnostic accuracy and clinical impact of MRI-based technologies for patients with non-alcoholic fatty liver disease: Systematic review and economic evaluation. *Health Technol. Assess* **2023**, *27*, 1–115. [CrossRef]

9. Kuroda, H.; Oguri, T.; Kamiyama, N.; Toyoda, H.; Yasuda, S.; Imajo, K.; Suzuki, Y.; Sugimoto, K.; Akita, T.; Tanaka, J.; et al. Multivariable Quantitative US Parameters for Assessing Hepatic Steatosis. *Radiology* **2023**, *309*, e230341. [CrossRef]

10. Paige, J.S.; Bernstein, G.S.; Heba, E.; Costa, E.A.C.; Fereirra, M.; Wolfson, T.; Gamst, A.C.; Valasek, M.A.; Lin, G.Y.; Han, A.; et al. A Pilot Comparative Study of Quantitative Ultrasound, Conventional Ultrasound, and MRI for Predicting Histology-Determined Steatosis Grade in Adult Nonalcoholic Fatty Liver Disease. *AJR Am. J. Roentgenol.* **2017**, *208*, W168–W177. [CrossRef] [PubMed]

11. Ozturk, A.; Kumar, V.; Pierce, T.T.; Li, Q.; Baikpour, M.; Rosado-Mendez, I.; Wang, M.; Guo, P.; Schoen, S.; Gu, Y.; et al. The Future Is Beyond Bright: The Evolving Role of Quantitative US for Fatty Liver Disease. *Radiology* **2023**, *309*, e223146. [CrossRef] [PubMed]

12. Kadi, D.; Loomba, R.; Bashir, M.R. Diagnosis and Monitoring of Nonalcoholic Steatohepatitis: Current State and Future Directions. *Radiology* **2024**, *310*, e222695. [CrossRef] [PubMed]

13. Şendur, H.N.; Özdemir Kalkan, D.; Cerit, M.N.; Kalkan, G.; Şendur, A.B.; Özhan Oktar, S. Hepatic Fat Quantification With Novel Ultrasound Based Techniques: A Diagnostic Performance Study Using Magnetic Resonance Imaging Proton Density Fat Fraction as Reference Standard. *Can. Assoc. Radiol. J.* **2023**, *74*, 362–369. [CrossRef] [PubMed]

14. Oh, N.; Kim, J.-H.; Rhu, J.; Jeong, W.K.; Choi, G.-s.; Kim, J.M.; Joh, J.-W. Automated 3D liver segmentation from hepatobiliary phase MRI for enhanced preoperative planning. *Sci. Rep.* **2023**, *13*, 17605. [CrossRef]

15. Ansari, M.Y.; Abdalla, A.; Ansari, M.Y.; Ansari, M.I.; Malluhi, B.; Mohanty, S.; Mishra, S.; Singh, S.S.; Abinahed, J.; Al-Ansari, A.; et al. Practical utility of liver segmentation methods in clinical surgeries and interventions. *BMC Med. Imaging* **2022**, *22*, 97. [CrossRef]

16. Senthilvelan, J.; Jamshidi, N. A pipeline for automated deep learning liver segmentation (PADLLS) from contrast enhanced CT exams. *Sci. Rep.* **2022**, *12*, 15794. [CrossRef] [PubMed]

17. Gu, Q.; Zhang, H.; Cai, R.; Sui, S.Y.; Wang, R. Segmentation of liver CT images based on weighted medical transformer model. *Sci. Rep.* **2024**, *14*, 9887. [CrossRef] [PubMed]

18. Gotra, A.; Sivakumaran, L.; Chartrand, G.; Vu, K.N.; Vandenbroucke-Menu, F.; Kauffmann, C.; Kadoury, S.; Gallix, B.; de Guise, J.A.; Tang, A. Liver segmentation: Indications, techniques and future directions. *Insights Imaging* **2017**, *8*, 377–392. [CrossRef]

19. Rahman, H.; Bukht, T.F.N.; Imran, A.; Tariq, J.; Tu, S.; Alzahrani, A. A Deep Learning Approach for Liver and Tumor Segmentation in CT Images Using ResUNet. *Bioengineering* **2022**, *9*, 368. [CrossRef]

20. Kumar, S.S.; Vinod Kumar, R.S. Literature survey on deep learning methods for liver segmentation from CT images: A comprehensive review. *Multimed. Tools Appl.* **2024**, *83*, 71833–71862. [CrossRef]

21.   Gross, M.; Huber, S.; Arora, S.; Ze'evi, T.; Haider, S.P.; Kucukkaya, A.S.; Iseke, S.; Kuhn, T.N.; Gebauer, B.; Michallek, F.; et al. Automated MRI liver segmentation for anatomical segmentation, liver volumetry, and the extraction of radiomics. *Eur. Radiol.* **2024**, *34*, 5056–5065. [CrossRef] [PubMed]

22.   Chlebus, G.; Schenk, A.; Moltz, J.H.; van Ginneken, B.; Hahn, H.K.; Meine, H. Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. *Sci. Rep.* **2018**, *8*, 15497. [CrossRef] [PubMed]

23.   Gul, S.; Khan, M.S.; Bibi, A.; Khandakar, A.; Ayari, M.A.; Chowdhury, M.E.H. Deep learning techniques for liver and liver tumor segmentation: A review. *Comput. Biol. Med.* **2022**, *147*, 105620. [CrossRef] [PubMed]

24.   Wu, J.; Liu, F.; Sun, W.; Liu, Z.; Hou, H.; Jiang, R.; Hu, H.; Ren, P.; Zhang, R.; Zhang, X. Boundary-aware convolutional attention network for liver segmentation in ultrasound images. *Sci. Rep.* **2024**, *14*, 21529. [CrossRef]

25.   Ali, A.-R.; Guo, P.; Samir, A. Liver Segmentation in Ultrasound Images Using Self-Supervised Learning with Physics-inspired Augmentation and Global-Local Refinement. In Proceedings of the Canadian Conference on Artificial Intelligence, Montreal, QC, Canada, 5–9 June 2023. [CrossRef]

26.   Ansari, M.Y.; Yang, Y.; Meher, P.K.; Dakua, S.P. Dense-PSP-UNet: A neural network for fast inference liver ultrasound segmentation. *Comput. Biol. Med.* **2023**, *153*, 106478. [CrossRef] [PubMed]

27.   Bhatia, V.; Hijioka, S.; Hara, K.; Mizuno, N.; Imaoka, H.; Yamao, K. Endoscopic ultrasound description of liver segmentation and anatomy. *Dig. Endosc.* **2014**, *26*, 482–490. [CrossRef]

28.   Esneault, S.; Hraiech, N.; Delabrousse, E.; Dillenseger, J.L. Graph cut liver segmentation for interstitial ultrasound therapy. In Proceedings of the 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Lyon, France, 22–26 August 2007; Volume 2007, pp. 5247–5250. [CrossRef]

29.   Ji, Z.; Che, H.; Yan, Y.; Wu, J. BAG-Net: A boundary detection and multiple attention-guided network for liver ultrasound image automatic segmentation in ultrasound guided surgery. *Phys. Med. Biol.* **2024**, *69*, 035015. [CrossRef]

30.   Zhang, L.; Wu, X.; Zhang, J.; Liu, Z.; Fan, Y.; Zheng, L.; Liu, P.; Song, H.; Lyu, G. SEG-LUS: A novel ultrasound segmentation method for liver and its accessory structures based on muti-head self-attention. *Comput. Med. Imaging Graph.* **2024**, *113*, 102338. [CrossRef]

31.   Gillies, D.J.; Rodgers, J.R.; Gyacskov, I.; Roy, P.; Kakani, N.; Cool, D.W.; Fenster, A. Deep learning segmentation of general interventional tools in two-dimensional ultrasound images. *Med. Phys.* **2020**, *47*, 4956–4970. [CrossRef]

32.   Song, K.D. Current status of deep learning applications in abdominal ultrasonography. *Ultrasonography* **2021**, *40*, 177–182. [CrossRef] [PubMed]

33.   Alves, V.P.V.; Dillman, J.R.; Tkach, J.A.; Bennett, P.S.; Xanthakos, S.A.; Trout, A.T. Comparison of Quantitative Liver US and MRI in Patients with Liver Disease. *Radiology* **2022**, *304*, 660–669. [CrossRef] [PubMed]

34.   Kang, H.Y.; Zhang, W.; Li, S.; Wang, X.; Sun, Y.; Sun, X.; Li, F.X.; Hou, C.; Lam, S.K.; Zheng, Y.P. A comprehensive benchmarking of a U-Net based model for midbrain auto-segmentation on transcranial sonography. *Comput. Methods Programs Biomed.* **2024**, *258*, 108494. [CrossRef] [PubMed]

35.   Cortacero, K.; McKenzie, B.; Müller, S.; Khazen, R.; Lafouresse, F.; Corsaut, G.; Van Acker, N.; Frenois, F.-X.; Lamant, L.; Meyer, N.; et al. Evolutionary design of explainable algorithms for biomedical image segmentation. *Nat. Commun.* **2023**, *14*, 7112. [CrossRef] [PubMed]

36.   Chollet, F. keras. 2015. Available online: https://github.com/fchollet/keras (accessed on 17 November 2024).

37.   Bilal; Pant, M.; Zaheer, H.; Garcia-Hernandez, L.; Abraham, A. Differential Evolution: A review of more than two decades of research. *Eng. Appl. Artif. Intell.* **2020**, *90*, 103479. [CrossRef]