Midline-Constrained Loss in the Anatomical Landmark Segmentation of 3D Liver Models

Abdul Karim Abbas, Aodhan Gallagher, Theodora Vraimakis, James Borgars, Ahmad Najmi Mohamad Shahir, Jibran Raja, Abhinav Ramakrishnan, and Sharib Ali*

School of Computer Science, University of Leeds, Leeds, UK
s.s.ali@leeds.ac.uk

Abstract. Anatomical landmark segmentation involves identifying specific points or regions within an anatomical structure and is integral to diagnostic processes and surgical guidance. This paper focuses on the segmentation of landmarks in 3D liver models in order to highlight key structures such as the falciform ligament and the liver ridge. The study is motivated by the need to support intraoperative laparoscopic registration tasks, aiming to enhance preoperative-to-intraoperative image fusion and thereby improving the localisation of tumours and vessels within the liver. As current practices typically rely on manual annotation, a process that is time-consuming and prone to human error when performed by less experienced operators, there is a clear need for a more efficient and reliable solution. Some recent works on landmark prediction in 3D liver models either over-predict or under-predict these landmarks. To overcome these challenges, we introduce a novel loss function that enforces geometric constraints by aligning segmentation predictions with a computed central anatomical midline. This strategy not only improves overall anatomical alignment but also ensures that the predictions remain thin and precise, reducing the occurrence of overly broad or misaligned outputs. This approach is utilised in conjunction with the PointNet++ architecture, trained on an extensive combined dataset composed of three smaller datasets, alongside the P2ILF challenge dataset, amounting to 300 unique 3D liver models in total. Our results indicate that our proposed solution forms a robust and precise approach, laying a solid foundation for future advances and feasibility in 3D-2D liver registration for intraoperative use. To allow reproducibility of this work, we have shared our code at: https://github.com/ARMADILLO-VISION/midline-loss

Keywords: Deep Learning · Midline Loss · 3D Landmark Segmentation · Laparoscopic Liver Surgery · PointNet++

1 Introduction

Anatomical 3D landmark segmentation is an essential task in medical image analysis, as it enables the precise localisation of anatomical structures within

* Corresponding author

3D models [17,16]. The importance of using 3D models via preoperative 3D rendering compared to conventional 2D scans for planning surgical strategies can improve understanding of liver anatomy and help in tissue-preserving surgery [15]. 3D landmark segmentation is particularly important in the context of preoperative-to-intraoperative image fusion [12,6,3]. For this reason, the accurate alignment of 3D models extracted from 3D CT/MRI scans acquired prior to surgery with the intraoperative 2D liver images allows for the understanding and location of important anatomical structures relative to the patient's anatomy during surgery [12,10,6,2,11].

To achieve a high degree of accuracy in such 3D-2D registration, it is essential to have an automated method for 3D liver landmark segmentation that can accurately predict the positions of key anatomical landmarks on unlabelled liver models [12,10]. The current practice of manual annotation of preoperative 3D liver models is time-consuming and prone to error [22], which motivates the development of a method that can automatically segment the key anatomical landmarks in liver models [12,20]. These landmarks can then be used for the registration process, as they provide reference points for aligning the 3D liver models with the 2D intraoperative images, which can also be done using segmentation of similar anatomical landmarks. In this way, 3D registration is performed automatically, however, less desirable manual and semi-automatic methods still exist for this task.

At present, a study conducted by Acidi et al. [1] shows that the most common reported method of registration is manual registration, with 11 studies proposing this approach as of 2023. This is followed by 9 studies conducted into semi-automatic approaches of registration [1]. There is a clear lack of studies into automatic methods of registration, with 3 studies reporting entirely automated approaches [26,23,14]. A large factor behind this is the lack of publicly available datasets with annotated 3D liver models [10,4], with many datasets containing a limited number of samples [1]. As a result, there is a need for a larger amount of data to be used in deep learning tasks in automated registration, which forms the main approach to our presented solution. Other literature suggests data synthesis as an alternative means to this problem [12].

The recent P2ILF challenge competition [3] curated and released the first comprehensive dataset in this domain with 11 patients, comprised of 9 patients for training and 2 for testing. The authors also performed an objective comparison of methods for 2D landmark segmentation, 3D landmark segmentation and preoperative-to-intraoperative fusion. A number of deep learning methods used by six participating teams were evaluated. For the 3D anatomical landmark segmentation, teams used MeshCNN [7], Graph Convolutional Network (GCN) [9], and PointNet-based architectures [18]. The results of these methods have shown promise, however, they display a high degree of variability with dependency on the quality of the training data. Issues such as under-predicting or over-predicting false positive regions were observed, indicating the need for further improvements in this area of research.

In this work, we propose an approach for anatomical 3D landmark segmentation that addresses these limitations. Our methodology enhances segmentation precision by training the PointNet++ architecture [19] on a large combined dataset of liver models, resulting in an improvement of the generalisation capabilities and robustness of the network across diverse patient livers. In addition, we introduce a novel midline loss function that addresses issues with over-segmenting, generating aligned and thin segmentation outputs resembling the ground truth data. To achieve this, the specialised loss function penalises deviations from a calculated central anatomical midline, promoting consistent and accurate landmark delineation while also discouraging overly large predictions, enforcing thin and anatomically plausible structures. Our approach demonstrates generalisability on unseen scenarios, surpassing baseline with conventional losses and the top-performing P2ILF challenge teams.

2 Methodology

The proposed method uses the PointNet++ neural network in conjunction with an innovative loss formula and is trained on a larger combined dataset, allowing for more accurate segmentation in unseen and unlabelled liver models, thus addressing current challenges in landmark prediction. It is this sourcing of increased data samples along with the implemented midline loss that forms the novelty of this approach. The basis of the loss function is detailed in the following methodology.

2.1 Midline Loss

To achieve precise and thin segmentation of anatomical structures, we propose a midline loss function ($\mathcal{L}_{\text{midline}}$) that explicitly incorporates geometric considerations. This loss is designed to align the segmentation predictions with a computed central anatomical midline derived from ground-truth labels, as well as penalising large deviations that result in overly broad or misaligned outputs.

Let $\mathcal{X}$ be the set of predicted 3D segmentation points for any given class. To estimate the central anatomical axis, we perform a weighted Principal Component Analysis on $\mathcal{X}$, using the segmentation probabilities as weights. This yields a principal direction vector v, which draws the best fit axis through the subset of predicted points $\mathcal{X}$ and naturally emphasises high confidence points while down weighting noisy outliers. We then sample along v to generate the set of candidate midline points $\mathcal{C}$.

Each candidate point $c_k \in \mathcal{C}$ is refined via a differentiable soft-snapping process. This process serves as a smooth projection mechanism, producing a refined point c_k^* using a softmax-weighted combination of all segmentation points:

$$c_k^* = \sum_{i=1}^{N} \alpha_i \, x_i, \quad \text{where} \quad \alpha_i = \frac{\exp\big(-\lambda \, \|x_i - c_k\|\big)}{\sum_{j=1}^{N} \exp\big(-\lambda \, \|x_j - c_k\|\big)}, \tag{1}$$

The set of refined points $\mathcal{S}_p$ defines the predicted midline.

To measure alignment between the predicted segmentation and the midline, each point $x_i \in \mathcal{X}$ is softly projected onto $\mathcal{S}_p$, producing a set of projected points $\hat{\mathcal{X}}$. These projections are subsequently used in the loss calculation to penalise deviations from the midline in a differentiable manner.

The midline loss consists of two complementary components: a deviation loss ($\mathcal{L}_{\text{dev}}$) and an alignment loss ($\mathcal{L}_{\text{align}}$). The deviation loss acts as a thickness controller by penalising the distance between each segmentation point and its nearest soft projection onto the computed midline. In doing so, it encourages the segmentation to remain narrow and closely bound to the midline, thereby reducing the overall thickness of the predicted structure. To control outliers, distances are clamped as needed. A visualisation of this process is shown in Fig. 1. Formally, the deviation loss is defined as:

$$\mathcal{L}_{\text{dev}} = \frac{\sum_{i=1}^{N} w_i \left(e^{\alpha \|x_i - \hat{x}_i\|} - 1 \right)}{\sum_{i=1}^{N} w_i + \epsilon}, \tag{2}$$

where x_i are the segmented points, $\hat{x}_i$ denote their differentiable projections onto the midline, w_i are the segmentation probabilities, and ϵ is a small constant to prevent division by zero.

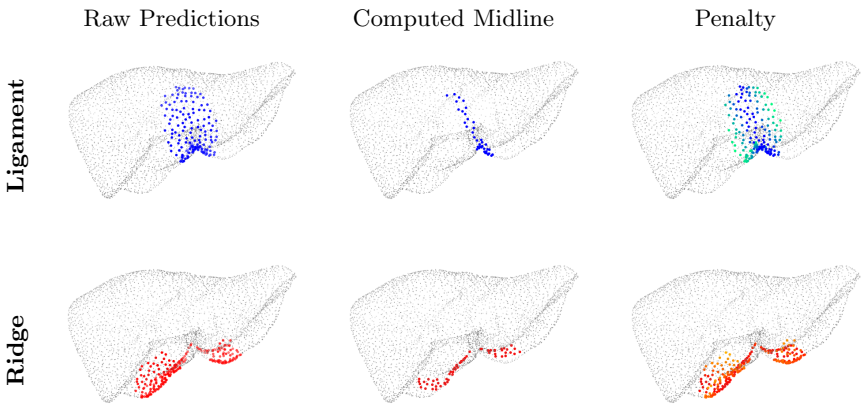

Fig. 1. Visualisation of the penalty computation pipeline for ligament and ridge regions. From left to right: raw segmentation predictions, computed midlines, and resulting deviation-based penalty maps. For ligament, penalty values range from low (blue) to high (green); for ridge, low penalties are shown in red, increasing to yellow for higher deviations from the midline.

In parallel, the alignment loss ($\mathcal{L}_{\text{align}}$) ensures that the predicted midline (S_p) aligns closely with the ground truth midline (S_{GT}). The predicted midline is derived using wPCA and soft-snapping, while the ground truth midline is

obtained by uniformly resampling the ground-truth annotated points. A soft-assigned Chamfer distance measures the spatial discrepancy between these midlines:

$$\mathcal{L}_{\text{align}} = \frac{1}{2}\Big(\frac{1}{|S_p|}\sum_{s_p \in S_p} d_{\text{CD}}(s_p, S_{\text{GT}}) + \frac{1}{|S_{\text{GT}}|}\sum_{s_{\text{GT}} \in S_{\text{GT}}} d_{\text{CD}}(s_{\text{GT}}, S_p)\Big), \tag{3}$$

where $d_{\text{CD}}(\cdot,\cdot)$ denotes the Chamfer point-to-midline distance.

The overall midline loss is then defined as:

$$\mathcal{L}_{\text{midline}} = (1-\lambda_m)\,\mathcal{L}_{\text{dev}} + \lambda_m\,\mathcal{L}_{\text{align}}, \tag{4}$$

with the hyper-parameter λ_m balancing the contributions of the deviation and alignment losses.

To balance the geometry-based midline loss with the weighted cross-entropy loss ($\mathcal{L}_{wCE}$), whose magnitudes may differ significantly during training, we introduce a dynamic scaling factor. This scaling factor ensures that both the geometric constraints and region-based accuracy are optimised simultaneously. The scale is computed as:

$$\beta = \frac{\mathcal{L}_{wCE}}{\frac{1}{2}\left(\mathcal{L}^{\text{ridge}}_{\text{midline}} + \mathcal{L}^{\text{lig}}_{\text{midline}}\right) + \epsilon}, \tag{5}$$

where $\mathcal{L}^{\text{ridge}}_{\text{midline}}$ and $\mathcal{L}^{\text{lig}}_{\text{midline}}$ each correspond to the same midline formulation ($\mathcal{L}_{\text{midline}}$) computed for the ridge and ligament structures separately. Finally, the combined total loss, incorporating weighted cross-entropy, is expressed as:

$$\mathcal{L}_{\text{total}} \;=\; \alpha\,\mathcal{L}_{wCE} \;+\; \beta\left(\lambda_{\text{ridge}}\,\mathcal{L}^{\text{ridge}}_{\text{midline}} \;+\; \lambda_{\text{lig}}\,\mathcal{L}^{\text{lig}}_{\text{midline}}\right). \tag{6}$$

2.2 PointNet++

The PointNet++ architecture used in our approach is designed to process and segment 3D point clouds directly, making the conversion of 3D data into point cloud representations a necessary step in our method. The implementation used is adapted from the repository by Yan [25] and is based on the work demonstrated by Qi et al [19]. The PointNet++ network was trained on the datasets using different loss functions, including the novel midline loss. The optimal hyperparameters and loss functions were selected through an ablation study, detailed in subsection 3.2. The standard PointNet architecture was also tested [18].

3 Experimental Setup

3.1 Dataset and Data Preparation

To ensure the robustness and generalisation of our proposed method, we used two publicly available datasets of 3D liver models, including a large dataset obtained from Zhang et al. [27] that combines three public datasets (3Dircadb [21],

LiTS [5], and Amos [8]). The other smaller dataset, consisting of 9 training and 2 test patients, was sourced from the P2ILF challenge [4]. Data for each patient includes a 3D liver model, saved as a wavefront object, and an XML file containing the anatomical annotations. These files are parsed to extract detailed contour information of the type of anatomical structure (ridge or ligament) and the corresponding indices of mesh vertices that define these contours. These extracted annotations were then used to create numerical labels for the 3D mesh vertices. Initially, all vertices are assigned a background label, which is subsequently updated to reflect the presence of specific anatomical features based on the parsed contour data.

To prepare the data for compatibility with the PointNet++ model, the 3D meshes were converted into point cloud arrays. These point clouds typically contain between 4000 and 15000 points, with each point inheriting the label corresponding to the anatomical structures previously identified in the mesh. Our conversion process ensures that each point cloud is standardised by either furthest point sampling or padding the vertices to achieve a fixed number of 4096 points.

Considering the P2ILF dataset contains only 11 patients in total, two patients are pre-emptively reserved for the test set. From the remaining 9, we adopt a 7–2 split for training and validation. This maintains a reasonable balance between training capacity and validation reliability, given the limited data, while preserving subject independence across all subsets. A similar ratio is used for consistency when applying the split to the combined dataset.

The data is normalised, ensuring all datasets share the same normalisation parameters. Class weights are computed using the inverse square root class frequency weighting method, assigning higher weights to classes with lower frequencies. This is necessary due to heavy class imbalance, with very low frequency in the ligament class compared to the liver class.

We apply augmentation techniques to increase the diversity of the training data and improve the model's generalisation capabilities, as well as its performance on unseen samples. Three different augmentations were performed, including upscaling, downscaling and rotations on the z-axis. Point clouds are randomly scaled in the range of 65%-145% of their original size, and then randomly rotated in the range of $-180°$-$180°$on the z-axis.

3.2 Ablation Study and Hyper-Parameter Selection

To systematically evaluate the impact of the learning rate and loss function components (Equation 6) on segmentation performance, a structured grid search was performed. Each configuration is trained independently, and the resulting segmentation quality is evaluated using 3D Chamfer distance [24]. Learning rate selection is performed separately through a separate ablation. The evaluated learning rates ranged from 0.00025 to 0.01, examining their effects on convergence stability and final segmentation accuracy. All experiments were conducted on an NVIDIA RTX 4070. A batch size of 32 was set and an AdamW optimiser with default parameters was utilised for training [13]. A learning rate scheduler

was implemented to reduce the learning rate once the validation loss plateaued and early stopping was applied if the validation loss remained stagnant.

In our experiments, negative log-likelihood (NLL) loss was adopted as the baseline due to its use as the standard loss in PointNet [18] and PointNet++ [19]. We also evaluated the weighted cross-entropy (wCE) loss, which incorporates the softmax operation internally. Although both losses yield similar outcomes when properly configured, the slight differences in their implementation can influence convergence behaviour and final accuracy, making the inclusion of wCE a valuable alternative for comparison.

The optimal hyper-parameters were selected based on performance on the validation dataset. For the combined dataset [27], the best-performing configuration used a weighted cross-entropy weight of $\alpha_{wce} = 0.25$, with geometry-based ridge and ligament losses set at $\lambda_{\text{ridge}} = 0.5$ and $\lambda_{\text{lig}} = 0.25$. A learning rate of 0.0075 was chosen, as it yields the lowest Chamfer distances. For the P2ILF dataset, the best configuration selected used $\alpha_{ce} = 0.25$, $\lambda_{\text{ridge}} = 0.75$, and $\lambda_{\text{lig}} = 0$, with an optimal learning rate of 0.0005.

4 Results

4.1 Evaluation Metrics

We assess the performance of our segmentation framework using 3D Chamfer distance, which quantifies the point-to-point average distance between the segmented landmarks and the ground truth [24]. It functions by averaging the minimum distance between points in two point clouds. For two point sets X and Y, it is defined as:

$$d_{\text{CD}}(X, Y) = \frac{1}{|X|} \sum_{x \in X} \min_{y \in Y} \|x - y\|_p + \frac{1}{|Y|} \sum_{y \in Y} \min_{x \in X} \|y - x\|_p, \tag{7}$$

where $\| \cdot \|_p$ denotes the p-norm distance of the points.

4.2 Quantitative Results

The quantitative results recorded during the series of ablations are presented in Table 1 and Table 2. Table 1 presents our findings when our models are tested on the P2ILF challenge dataset [4] and Table 2 shows our findings tested on the combined dataset [27]. Furthermore, we include the results of the top two performing teams from the P2ILF challenge paper as a means of comparison against other literature [4]. The two teams included from the P2ILF challenge are: UCL, which achieved the best results on the liver ligament, and NCT, which achieved the best results on the liver ridge. Each model is tested on the hold-out set of the dataset it was trained on, as well as the other dataset to verify generalisation capabilities.

Table 2 indicates that when testing on the combined dataset, the two best performing configurations of PointNet++ with weighted cross-entropy loss and

Table 1. Evaluation on the P2ILF challenge test set [4]. LR: learning rate. *ch_r*, *ch_l*: Chamfer distances (in mm) for ridge and ligament. ‡: trained on combined dataset [27]. *: evaluation on unseen test set. Highlighted cells outperform all P2ILF teams; best results are bold.

	Model	**Loss**	**LR**	**Train**	**ch_r**	**ch_l**	**Mean**
	PointNet++	Midline,wCE	0.01	‡*	**19.70**	**13.46**	**16.58**
	PointNet++	NLL	0.0005	P2ILF	22.23	55.77	39.00
	PointNet++	wCE	0.01	‡*	23.72	60.26	41.99
	PointNet++	wCE	0.0005	P2ILF	38.43	69.62	54.03
	PointNet++	Midline,wCE	0.00075	P2ILF	36.11	116.10	76.11
	PointNet	NLL	0.01	P2ILF	95.85	73.46	84.65
	PointNet++	NLL	0.005	‡*	F	115.54	115.54
	PointNet	NLL	0.00075	‡*	199.94	F	199.94
Teams [3]							
UCL	PointNet++	NLL,HFD	0.001	P2ILF	27.97	24.47	26.22
NCT	2×MeshCNN	wCE	–	P2ILF	27.19	36.38	31.79

Table 2. Evaluation on the combined test set [27]. LR: learning rate. *ch_r*, *ch_l*: Chamfer distances (in mm) for ridge and ligament. ‡: trained on combined dataset. *: evaluation on unseen test set. Best results are bold.

Model	**Loss**	**LR**	**Train**	**ch_r**	**ch_l**	**Mean**
PointNet	NLL	0.00075	‡	7.92	**16.62**	**12.27**
PointNet++	wCE	0.01	‡	**7.13**	18.30	12.71
PointNet++	Midline,wCE	0.01	‡	11.62	21.57	16.60
PointNet++	Midline,wCE	0.00075	P2ILF*	19.50	23.21	21.35
PointNet++	NLL	0.0005	P2ILF*	28.79	17.95	23.37
PointNet	NLL	0.01	P2ILF*	27.39	29.10	28.25
PointNet++	wCE	0.0005	P2ILF*	33.49	23.48	28.49
PointNet++	NLL	0.005	‡	32.21	30.39	31.30

PointNet with negative loss likelihood loss achieve the lowest mean Chamfer distance and ultimately outperform the best performing configurations from the P2ILF challenge [4]. However, it should be noted that despite the high performance on the combined dataset, these configurations generalise poorly to the P2ILF dataset, which can be observed in Table 1. For instance, the PointNet++ configuration using weighted cross-entropy loss records a mean Chamfer distance of $12.71mm$ on the combined dataset but deteriorates sharply to $41.99mm$ when generalising to the P2ILF dataset. Similarly, the PointNet model with NLL loss (LR = 0.00075) achieves a competitive $12.27mm$ on the combined dataset, yet it fails to generalise, reaching $199.94mm$ on the P2ILF dataset.

Furthermore, Table 1 highlights that PointNet++ with combined midline and weighted cross-entropy loss functions along with a learning rate of 0.01, when trained on the combined dataset and tested on the P2ILF dataset, outperforms all teams from the P2ILF challenge and achieves the best results out of all our models tested on the P2ILF data. Although the Chamfer distances reported by this model are lower than those presented in Table 2, they nonetheless demonstrate that our proposed method generalises effectively to unseen data, as evidenced by its performance on the hold-out set. For example, this method, using the midline loss at LR = 0.01, reduces the mean Chamfer distance to $16.58mm$ – a reduction exceeding 36% compared to the UCL team's method.

4.3 Qualitative Results

Figure 2 provides a comparative visual analysis of segmentation outputs from models trained on the combined dataset for four patients. Two from the P2ILF test set (patient 4 and patient 11) [4] and two from the combined dataset (LiTS-65 and Amos-119) [27]. Figure 3 is also included as a comparison between the top performing teams in the P2ILF challenge and our best model.

The results indicate that the segmentation produced with the midline loss is more closely bound to the ground truth. Notably, for both patient 4 and patient 11 in Figure 2, the ridge structure is more accurately delineated as the predicted midline closely curves around the ridge, which contrasts the broader, region-like segmentations observed with the weighted cross-entropy model. Although the overall segmentation size remains relatively large for the ligament, the localisation is notably improved with the midline loss, showing a more generalised and anatomically consistent alignment.

These improved results are also evident in the comparisons made in Figure 3. Segmentations produced by our model demonstrate improvements over the two highest scoring P2ILF teams. In particular, the predicted ridge and ligament regions are more accurately aligned with the ground truth in contrast to the top two P2ILF team's predictions, which are erratic and have incorrectly placed some points behind the liver.

By incorporating a geometric aspect, the midline loss improves localisation by effectively carving the ridge around the liver while ensuring thin segmentation. Similar improvements in localisation and accuracy are also observed on the combined dataset patients (LiTS-65 and Amos-119), particularly for ridge segmentations.

5 Discussion

Our experimental results provide compelling quantitative evidence for the efficacy of our proposed method in 3D liver landmark segmentation. By leveraging a significantly larger combined dataset of 300 unique 3D liver models [27] and incorporating our geometry-constrained midline loss, our PointNet++ model achieves mean Chamfer distances of $16.60mm$ on the combined dataset and

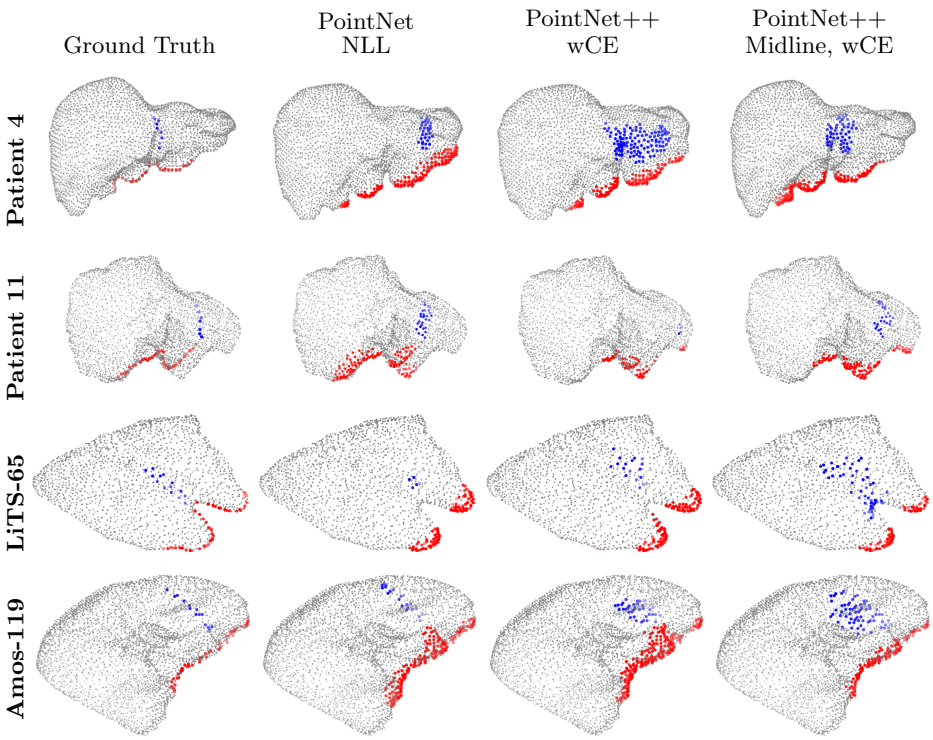

Fig. 2. Qualitative comparison of segmentation results between our proposed method on baselines from P2ILF [4] (patients 4 and 11) and the combined dataset [27] (LiTS-65 [5] and Amos-119 [8]). Ligament points are coloured blue; Ridge points are coloured red.

16.58mm on the P2ILF dataset. These outcomes contrast markedly with previously seen approaches, underscoring the limitations faced by earlier methods, particularly those from the P2ILF challenge teams which were constrained by limited training data and consequently struggled to generalise. In contrast, our approach benefits from both a larger, more diverse dataset and a novel loss function that explicitly enforces geometric consistency by penalising deviations from a computed central anatomical midline. This strategy promotes thin and greater-aligned segmentations that more accurately reflect true anatomical structures. Moreover, the enhanced performance is attributed to the incorporation of geometric constraints, which provides stronger regularisation and enables the model to capture complex anatomical variations more robustly. The best results observed on the combined dataset [27], even when using a P2ILF-trained model, highlight the substantial improvements in generalisation achieved by our approach. In particular, the incorporation of the midline loss has proven critical for enabling robust predictions across diverse liver datasets. As shown in Table 2, even when the PointNet++ model with midline and weighted cross-entropy

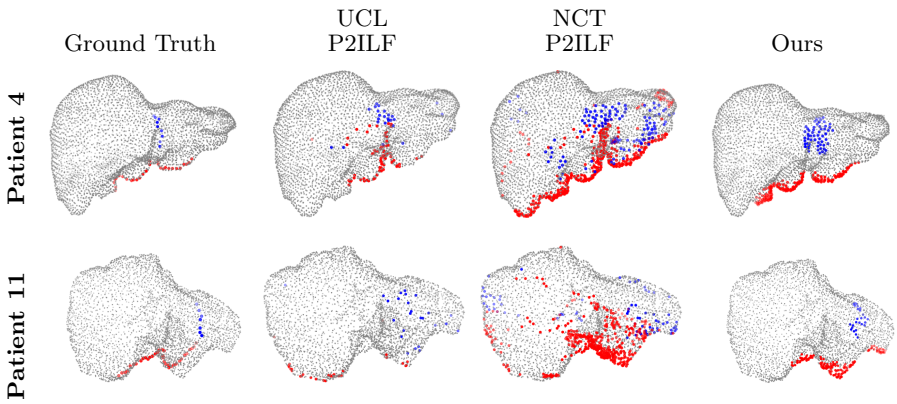

Fig. 3. Qualitative comparison of segmentation results between the two best teams from the P2ILF challenge [4] and our best model. Test patients 4 and 11 are used. Ligament points are coloured blue; Ridge points are coloured red. Points on the backside of the liver are displayed with a lower alpha value.

loss was trained solely on the smaller P2ILF dataset (with a learning rate of 0.00075), it still performed relatively well on the combined dataset, achieving a mean Chamfer distance of 21.35mm compared to 16.60mm for the model trained on the larger combined dataset (with a learning rate of 0.01).

Furthermore, visual comparisons reveal that the segmentation outputs using the midline loss exhibit a distinct, precise curvature of the ridge that is more tightly bound to the ground truth. In contrast, predictions from models using solely wCE tend to be broader and less defined, resembling a region-like segmentation. The improved localisation achieved by effectively demarcating the ridge around the liver not only confirms the enhanced geometric fidelity of our method but also demonstrates its ability to generalise well to unseen data from different datasets. However, in the particularly challenging LITS 65 case, the ridge is deformed enough that it differs significantly from its typical anatomy and all models detect the two segments separately rather than one continuous prediction. These mispredictions underscore that even other loss functions struggle with extreme anatomical outliers, highlighting the need for strategies specifically designed to handle such extreme cases. Regardless, there is improvement in cross-dataset performance and this is particularly evident when comparing the performance on the combined dataset to that of the P2ILF dataset, where our method substantially reduces mean Chamfer distances and outperforms previous configurations. Overall, the combination of a robust, extensive training dataset with a geometry-aware loss function leads to improvements in both quantitative metrics and qualitative visual assessments, establishing a strong foundation for future advancements in 3D liver segmentation.

Importantly, the enhanced landmark segmentation accuracy carries direct benefits for the laparoscopic 3D–2D registration pipeline, especially when considering typical clinical workflows, whereby preoperative 3D liver models must

be aligned with intraoperative laparoscopic images to guide instrument navigation and tumour resection [12,10,6]. More precise 3D landmark localisation reduces the spatial uncertainty in the initial alignment step, leading to lower registration error and more stable convergence of feature-based registration algorithms [12,3]. In practice, this can translate to faster registration times and improved spatial overlay.

Despite these advances, a significant limitation remains in the relatively small number of annotated ligament landmarks. The lack of ground-truth labels makes it inherently challenging for the model to learn precise geometry. This scarcity likely contributes to the broad localisation predictions observed in our qualitative results, although the midline loss does mitigate this issue by enforcing geometric consistency to an extent. In practice, the lack of fine-grained detail in the training labels limits the model's ability to capture subtle variations in ligament morphology, regardless of the loss function used. Addressing this will require more extensive high-resolution annotations or targeted data augmentation strategies.

6 Conclusion

In this work, we presented a novel approach for 3D liver segmentation that enhances anatomical accuracy by integrating a geometry-aware midline loss function using a larger, combined dataset. Our method yields notably more precise and robust segmentations than those produced by traditional loss functions, effectively addressing common issues such as overly thick or misaligned predictions observed in previous works. By explicitly constraining the segmentation with geometric information, our approach achieves improved alignment with true anatomical structures and demonstrates superior generalisation across diverse datasets. A major limitation remains in the quality and quantity of existing data. As such, future research should focus on expanding training datasets to encompass greater anatomical diversity, integrating a geometrically constrained loss function that is more tolerant to outliers and the investigation of geometric recognition features within model architectures.

Acknowledgments. This project is part of the National Institute for Health and Care Research (NIHR) Leeds Biomedical Research Centre (BRC) (NIHR203331) pump-priming funding "ARMADILLO" project. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Acidi, B., Ghallab, M., Cotin, S., Vibert, E., Golse, N.: Augmented reality in liver surgery. Journal of Visceral Surgery **160**(2), 118–126 (2023). https://doi.org/10.1016/j.jviscsurg.2023.01.008
2. Adagolodjo, Y., Trivisonne, R., Haouchine, N., Cotin, S., Courtecuisse, H.: Silhouette-based pose estimation for deformable organs: Application to surgical augmented reality. In: Proc. IEEE/RSJ IROS. pp. 539–544. IEEE (2017). https://doi.org/10.1109/IROS.2017.8202205
3. Ali, S., Espinel, Y., Jin, Y., Liu, P., Güttner, B., Zhang, X., et al.: An objective comparison of methods for augmented reality in laparoscopic liver resection by preoperative-to-intraoperative image fusion. CoRR **abs/2401.15753** (2024), arXiv:2401.15753
4. Ali, S., Jin, Y., Lopez, Y.E., Bartoli, A.: Preoperative to intra-operative laparoscopic fusion challenge (p2ilf) (2022). https://doi.org/10.5281/zenodo.6362161, dataset and challenge website, accessed 24 May 2025
5. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., et al.: The liver tumour segmentation benchmark (lits). Medical Image Analysis **84**, 102680 (2023). https://doi.org/10.1016/j.media.2022.102680
6. Espinel, Y., Calvet, L., Botros, K., Buc, E., Tilmant, C., Bartoli, A.: Using multiple images and contours for deformable 3d–2d registration of a pre-operative ct in laparoscopic liver surgery. International Journal of Computer Assisted Radiology and Surgery **17**(12), 2211–2219 (2022). https://doi.org/10.1007/s11548-022-02774-1
7. Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S., Cohen-Or, D.: Meshcnn: A network with an edge. CoRR **abs/1809.05910** (2018), arXiv:1809.05910
8. Ji, Y., Bai, H., Ge, C., Yang, J., et al.: Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. In: Advances in Neural Information Processing Systems. vol. 35, pp. 36722–36732 (2022)
9. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. CoRR **abs/1609.02907** (2016), arXiv:1609.02907
10. Koo, B., Robu, M.R., Allam, M., Pfeiffer, M., Thompson, S., et al.: Automatic, global registration in laparoscopic liver surgery. International Journal of Computer Assisted Radiology and Surgery **17**(1), 167–176 (2022). https://doi.org/10.1007/s11548-021-02518-7
11. Koo, B., Özgür, E., Roy, B.L., Buc, E., Bartoli, A.: Deformable registration of a preoperative 3d liver volume to a laparoscopy image using contour and shading cues. In: Lecture Notes in Computer Science 10435. pp. 326–334. Springer (2017). https://doi.org/10.1007/978-3-319-66182-7_38
12. Labrunie, M., Ribeiro, M., Mourthadhoi, F., Tilmant, C., Roy, B.L., Buc, E., Bartoli, A.: Automatic preoperative 3d model registration in laparoscopic liver resection. International Journal of Computer Assisted Radiology and Surgery **17**(8), 1429–1436 (2022). https://doi.org/10.1007/s11548-022-02641-z
13. Loshchilov, I., Hutter, F.: Decoupled weight decay regularisation. In: Proc. International Conference on Learning Representations (2019)
14. Luo, H., Yin, D., Zhang, S., Xiao, D., et al.: Augmented reality navigation for liver resection with a stereoscopic laparoscope. Computer Methods and Programs in Biomedicine **187**, 105099 (2020). https://doi.org/10.1016/j.cmpb.2019.105099
15. Montalti, R., Rompianesi, G., Cassese, G., Pegoraro, F., Giglio, M.C., De Simone, G., et al.: Role of preoperative 3d rendering for minimally

invasive parenchyma-sparing liver resections. HPB **25**(8), 915–923 (2023). https://doi.org/10.1016/j.hpb.2023.04.008
16. 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. Scientific Reports **13**, 17605 (2023). https://doi.org/10.1038/s41598-023-44736-w
17. Pei, J., Cui, R., Li, Y., Si, W., Qin, J., Heng, P.A.: Depth-driven geometric prompt learning for laparoscopic liver landmark detection. CoRR **abs/2406.17858** (2024), arXiv:2406.17858
18. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. CoRR **abs/1612.00593** (2017), arXiv:1612.00593
19. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. CoRR **abs/1706.02413** (2017), arXiv:1706.02413
20. Schneider, C., Thompson, S., Totz, J., Song, Y., Sodergren, M.H., et al.: Comparison of manual and semi-automatic registration in augmented-reality image-guided liver surgery: A clinical feasibility study. Surgical Endoscopy **34**(10), 4702–4711 (2020). https://doi.org/10.1007/s00464-020-07807-x
21. Soler, L., Hostettler, A., Agnus, V., Charnoz, A., et al.: 3D image reconstruction for comparison of algorithm database: A patient-specific anatomical and medical image database. Tech. rep., IRCAD, Strasbourg, France (2010)
22. Teatini, A., Pelanis, E., Aghayan, D., et al.: The effect of intraoperative imaging on surgical navigation for laparoscopic liver resection surgery. Scientific Reports **9**, 18687 (2019). https://doi.org/10.1038/s41598-019-54915-3
23. Thompson, S., Schneider, C., Bosi, M., Gurusamy, K., et al.: In vivo estimation of target registration errors during augmented-reality laparoscopic surgery. International Journal of Computer Assisted Radiology and Surgery **13**(6), 865–874 (2018). https://doi.org/10.1007/s11548-018-1761-3
24. Wu, T., Pan, L., Zhang, J., Wang, T., Liu, Z., Lin, D.: Balanced chamfer distance as a comprehensive metric for point cloud completion. In: Advances in Neural Information Processing Systems. vol. 34, pp. 29088–29100 (2021)
25. Yan, X.: Pointnet/pointnet++ in pytorch. https://github.com/yanx27/Pointnet_Pointnet2_pytorch (2019), gitHub repository, accessed 24 May 2025
26. Zhang, P., Luo, H., Zhu, W., Yang, J., et al.: Real-time navigation for laparoscopic hepatectomy using image fusion of pre-operative 3d surgical plan and intraoperative indocyanine green fluorescence imaging. Surgical Endoscopy **34**(8), 3449–3459 (2020). https://doi.org/10.1007/s00464-019-07121-1
27. Zhang, X., Ali, S., Han, M., Kang, Y., Wang, X., Zhang, L.: Two-stream meshcnn for key anatomical segmentation on the liver surface. International Journal of Computer Assisted Radiology and Surgery (2025), in press