Automatic Deep Deformable Registration using Domain Adaptation and Run-Time Optimisation

Emilien Gadoux[1,*][0009−0008−5357−6069] and Adrien Bartoli[1][0000−0003−3545−7329]

[1] Institut Pascal, Clermont-Ferrand, France
[*]Corresponding author: emilien.gadoux@uca.fr

Abstract. Augmented reality from preoperative 3D model registration is promising to assist navigation in minimally-invasive liver surgery. The current registration methods are either accurate, but require surgeon interactions to annotate anatomical landmarks, or are fully automatic, but inaccurate. We propose a two-step automatic and accurate registration method. Step 1) segments the registration landmarks with a neural method. Step 2) estimates the 3D model deformation from the landmarks. The task is challenging because of the defects of the automatically segmented landmarks and the impossibility to label registration for training. We handle it by combining supervised training from synthetic transformations with domain adaptation and a novel robust Run-Time Optimisation (RTO). Our method outperforms existing ones, both with manual and automatic landmark segmentations, improving both automation and accuracy.

Keywords: MIS · Liver · Registration · Domain Adaptation

1 Introduction

Minimally-invasive liver surgery offers significant advantages, but accurately localising intra-parenchymatous structures is highly challenging. Augmented reality is an appealing solution, where the patient's liver is modelled before surgery from a CT scan, and is then projected onto the surgical 2D image to reveal tumour locations. A key step is thus to register the preoperative 3D model to the surgical image, which is highly challenging, as the liver is extremely deformable and only partially visible during surgery. All liver registration methods exploit anatomical landmarks visible in both modalities. The vast majority of methods, whether based on numerical optimisation or neural models, and whether estimating a rigid [1, 9] or a deformable [11, 13] registration, rely on the availability of clean landmarks annotated by the surgeon on the surgical image. While some methods are accurate, with a Target Registration Error (TRE) within 10 to 20 mm, the requirement for surgeon interactions is problematic. Recent methods [10, 2] automatically segment the landmarks but their registration accuracy, with TREs of 20 to 45 mm, is insufficient for clinical use.

We propose a deep fully-automatic method reaching the level of accuracy of the manual methods. It is thus compatible with clinical use, both in terms

of user experience and surgical precision.Two major challenges are the unavailability of clinical data hampering mere supervised training of a neural network, and the fact that the patient's liver appearance only becomes available at the time of surgery. This is addressed by step 1), which implements patient-generic automatic, hence imperfect, landmark segmentation, and step 2), which implements landmark-based patient-specific deformation estimation. Our contributions lie in step 2), which we propose to split in two substeps as described in figure 1. Sub-step 2.A) estimates a global registration, modelled by a 3D affine transformation. This transformation is regressed by a neural network trained preoperatively from synthetic data, obtained by simulating the surgical camera pose to synthesise the segmented landmarks. A major hurdle is to cope with the domain gap between the perfectly synthesised landmarks and the automatically segmented landmarks on surgical images. We deal with this hurdle using Domain Adaptation (DA). Substep 2.B) estimates the residual registration deformation through a proposed Run-Time Optimisation (RTO) process. RTO uses a neural shape parameterisation and decodes the liver deformation. It limits the number of parameters to estimate and ensures that the deformation is physically plausible. It is trained using an auto-encoder, from deformations simulated by the Finite Element Method. RTO searches for the latent code by minimising the differences between the predicted and segmented landmarks. The proposed method named ADeLiR (Automatic Deformable Liver Registration) is accurate, obtaining a TRE of 13.60 mm on the standard benchmark, making it the first automatic and sufficiently accurate method for clinical applications.

2 Related work

Methods for 3D-2D registration generally have two parts: registration, to which our work belongs, and tracking. Early liver registration methods [17] relied on manual initialisation and manual landmarks. Deformation was addressed in [1, 9] by coarse initialisation [14] and refinement [11, 13]. These methods are sufficiently accurate for the clinical setting but require clean manual landmarks. The second-generation methods perform both automatic landmark segmentation and deformation estimation via neural networks to enable full automation [10, 2]. The most advanced method [10] uses a Liver Mesh Recovery (LMR) network for deformable registration. These automatic methods are still far from the manual methods in accuracy. This owes to the scarcity of the available datasets and the use of synthetic data ([10, 13] and several methods in [2]) being hindered by the domain gap. DA aims at reducing the domain gap, for a neural network trained on a source domain and tested on a different target domain [12]. DA generally forces the network to encode the source and target images with the same distribution, by adding a discrepancy loss term [19] or an adversarial network [5]. These techniques were used in medical image processing [6] for classification and detection. Recent DA methods [20] combine transformers and an adversarial network, reaching state-of-the-art performance in classification. We adapt these methods to regression for liver registration.

3 System and Methods

3.1 System Overview and Main Steps

The preoperative 3D model is a mesh model of the complete liver annotated with anatomical landmarks and its internal structures, which are only used at the ultimate augmented reality visualisation stage. We assume that the endoscope is calibrated, with intrinsic matrix K_{test}, and the surgical image corrected for optical distortions. The landmarks are anatomical edges of the liver [11]. In contrast to existing work, we do not require clean anatomical landmarks in the surgical image. We propose a fully-automatic method in two steps to estimate a 3D-3D geometric transformation G mapping the preoperative 3D model to the precise location where they were seen by the camera during surgery. Step 1) automatically segments the landmarks in the surgical image with an existing patient-generic model [11] trained for default intrinsic parameters K_{trained}. We gain independence to the camera parameters at test time by transforming the landmarks with $K_{\text{trained}}K_{\text{test}}^{-1}$. Step 2) estimates G from the noisy landmarks with a patient-specific neural model. Our contributions are in step 2), which has two sub-steps (figure 1): coarse registration, sub-step 2.A), estimates a 3D affine transformation with 12 parameters, and refinement, sub-step 2.B), estimates the residual 3D deformation with m parameters.

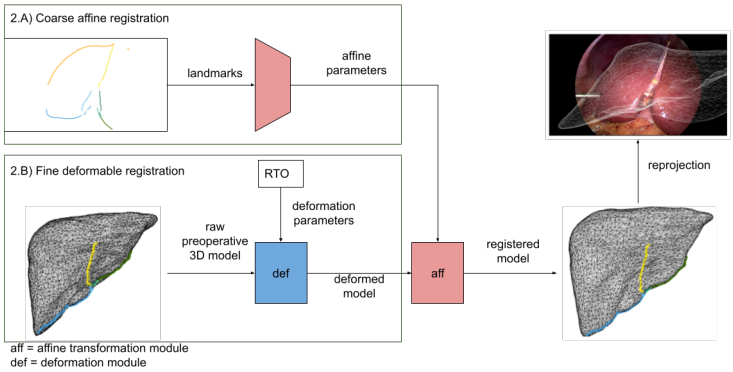

Fig. 1. Following automatic landmark segmentation step 1), the proposed method forming step 2) involves coarse step 2.A), inference of a 3D affine transformation, and fine step 2.B), inference of a deformation in the preoperative 3D model coordinate frame, to fit the anatomical landmarks.

3.2 Coarse Affine Registration

Affine registration captures the coarse global changes between the preoperative 3D model and the surgical camera with a 12 parameter 3D affine transformation.

Importantly, it has a linear parameterisation, escaping the orthonormal rotation constraints of the pose. The training data are clean simulated landmark images. However, our objective is to use automatic landmarks (with missing parts and perturbed points), inducing a domain gap, which we tackle with DA (figure 2).
General regression architecture. The network maps the landmarks to the 3D affine transformation. The input is an image whose channels contain the landmark masks, which handles missing landmarks. The outputs are the 12 affine parameters. The network encodes the landmark image and uses 2 fully-connected layers as decoder. We have experimented with encoders including ResNet and ViT. The affine transformation is then applied to the 3D model to obtain $\hat{y}_i$. The 3D loss L_{task} is the Mean Square Error MSE $= \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ with the simulated deformed model y_i. We chose this 3D loss over the classical reprojection loss as it better captures the 3D information. It can be trivially computed because the training data are simulated, as described in section 4.1.
Domain Adaptation. We train the model with DA to handle automatically-segmented landmarks. As we do not have corresponding cross-domain (corresponding simulated and segmented) data, we constrain the model to encode images from both domains with similar feature distributions. DA is trained with a supervised loss for the simulated landmarks and an adversarial loss for the real landmarks, referred to as training and test in section 2. DA training is specific to each segmentation method. We adapt two state-of-the-art DA methods, namely DANN [5] and TVT [20], to the regression task at hand. Both methods take synthetic and automatic image features as input and pass them through an adversarial decoder, trained to distinguish the source and target features with a binary Cross Entropy loss L_{adv}. A Gradient Reversal Layer [8] reverses the backpropagated gradient, training the encoder to fool the adversarial decoder by encoding the synthetic and automatic landmarks similarly. For DANN, we add its adversarial decoder and loss to our method (Resnet [7] and ViT [4] backbones) with final loss $L = L_{\text{task}} + L_{\text{adv}}$. For TVT, we adapt its architecture to regression by adding a linear head layer with 12 outputs for the sought affine parameters and change the Cross Entropy task loss to the MSE, keeping the adversarial term.

3.3 Fine Deformable Registration

The liver is a soft organ. Modelling its deformations is thus extremely important to achieve precise registration. Learning-based deformation modelling and estimation have shown potential for computed-assisted surgery [3], with neural networks trained to reproduce liver deformations simulated by the Finite Element Method [21]. We propose to refine the coarse registration by estimating the liver deformation, for which we introduce a latent-space representation and the RTO process, as shown in figure 3.
Deformed shape neural representation. We propose to use a latent-space neural representation to parameterise liver deformation as $D(z)$. Concretely, we train an 8 linear layers auto-encoder to reconstruct liver deformations, with 100 epochs, the Adam optimiser and a learning rate of 10^{-4}. A latent code

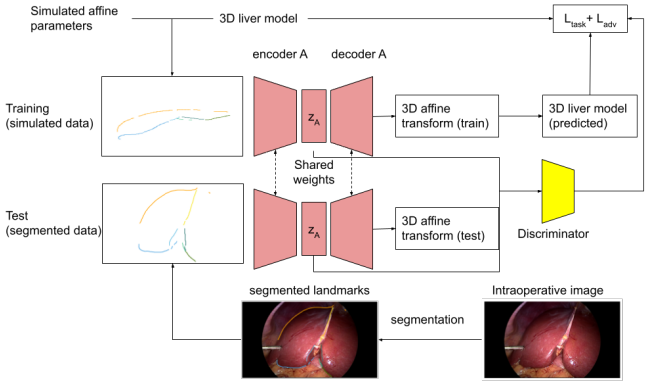

Fig. 2. The coarse registration step 2.A) uses DA. Encoder A and decoder A are trained on source images. DA is achieved by an adversarial network, leading to the discriminator, trained to distinguish the source from the target features while training the encoder A to fool the discriminator.

$z \in R^{256}$ can then be decoded by D to generate a deformed 3D model. The 3D model is represented by a mesh with approximately $9,000$ vertices stacked in $D(z)$. The training deformations are synthetically-generated by the Finite Element Method, as described in section 4.1. We simulate the deformations in a coordinate frame centred on the preoperative 3D model, fixing the inferior vena cava to create boundary constraints. This means that the parameterisation does not learn global transformations. We finally compose the produced deformation with the affine transformation, which at this point is held fixed.

Run-Time Optimisation. RTO deforms the liver model by optimising the latent code as $z = \min_z d_H(\Pi(D(z)), L)$ where Π is the combined affine transformation and projection function and L are the target landmarks. Function d_H sums the Hausdorff distance over the visible landmarks. The minimisation is implemented within the MeshSDF framework [16] with $z = 0$ as initial condition. As shown in figure 3b, the code z is decoded to a deformed liver mesh, transformed with the fixed affine transformation, reprojected using the intrinsic camera parameters, and the landmarks it carries are compared to the segmented ones via the Hausdorff distance. The gradient is then backpropagated to z.

4 Datasets

4.1 Synthetic Training and Validation Data

The two sub-steps of our method need to be preoperatively trained from synthetic data. Both trainings are patient-specific and do not require intraoperative training, which is compatible with the surgical workflow.

Surgical camera simulation. Training the model for affine transformations requires generating the image landmarks, hence to project the 3D model to form

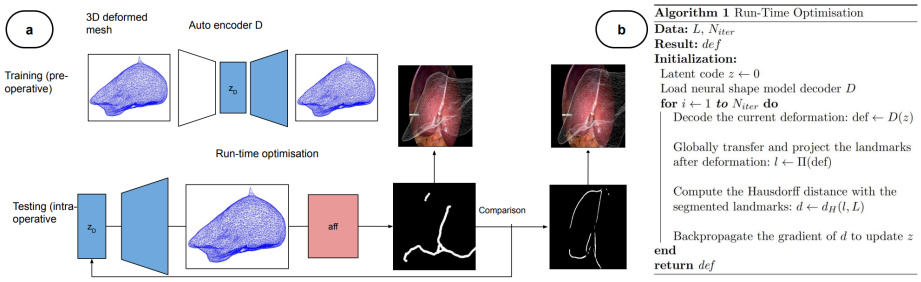

Fig. 3. Run-Time Optimisation (RTO). RTO is the fine registration step 2.B). It optimises the model landmark reprojection under guidance of a latent-space neural deformation model, trained preoperatively using an autoencoder and simulated deformations. a) RTO workflow, b) RTO algorithm.

2D images. We simulate rigid transformations representing camera pose as described next. The intrinsic parameters of the real surgical camera are however unknown when training. The proposed camera-generic training uses typical surgical camera parameters instead, referred to as the K_{trained} parameters. Specifically, we use a standard image resolution of 1920×1080 pixels, a focal length experimentally calibrated to $f = 992$ pixels with a checkerboard for the typical settings of a Karl Storz laparoscope with 0-degree optics EUP-OL334, and the principal point set at the image centre. Using standard camera parameters, we projected the landmarks annotated on the patient's 3D model and stored the resulting 2D projections for training. The proposed method reaches an average 3D registration error of 7.35 mm on the evaluation simulations.

Deformation simulation. We trained the autoencoder D using 2,000 simulated deformations in FEBio 1.6, a Finite Element Method software. We simulated tool forces exerted on the liver as surface forces, allowing the capture of a wide range of realistic deformations. We modelled the liver with an Ogden hyperelastic model [18] with a material density of $1,000$ kg/m^3 (comparable to water), a bulk modulus of 0.4 GPa (near-incompressibility) and standard values for the other model parameters ($c_1 = 4,100$ and $m_1 = 3.17$).

4.2 Real Test Data

We use the standard public benchmark [15], giving the liver and tumour 3D models for 4 patients and between 6 and 20 corresponding intraoperative images per patient. The tumour GT position, as measured with an ultrasound probe during surgery, allows one to compute the tumour TRE for the registration methods. The dataset includes clean manually-annotated landmarks but we also run automatic landmark segmentation within the proposed fully-automatic method. Whilst this dataset is interesting owing to the available GT, it has a small size and uses images where the LUS probe is visible and equipped with markers, which may perturb landmark segmentation. A known issue occurs for Patient 2,

whose liver undergone a major torsion during image acquisition. We thus give statistics with and without Patient 2. We also include 10 patients without ground truth, numbered from 11 to 20, to visualise the ADeLiR results. We compared the automatic to the manual landmarks, revealing an average point-noise level of 7.6 pixels.

5 Experimental Results in Training and Testing

All trainings were performed using an Nvidia RTX Ti 2080 GPU, except for TVT which used an Nvidia 4090 GPU. The coarse registration models were trained for 100 epochs, with 64 as batch size, 0.0005 as learning rate and the Adam optimiser, except for TVT which used SGD. The deformation autoencoders were trained similarly, with learning rate 10^{-4}. The proposed method is referred to

Method	Opt		LMR		ADeLiR (ResNet)		ADeLiR (ViT)		ADeLiR (TVT)	
Segmentation	M	A	M	A	M*	A	M*	A	M	A
Patient 1	10.3	29.3	17.4	17.9	**06.48**	**12.45**	26.46	29.34	45.07	27.63
Patient 2	63.0	86.9	53.8	45.6	43.57	46.44	74.09	70.14	**42.59**	**45.01**
Patient 3	**09.5**	30.0	17.6	46.9	13.46	**17.13**	54.40	63.45	27.02	30.51
Patient 4	14.7	19.8	17.0	22.7	**10.18**	**11.20**	40.49	39.04	101.86	105.87
Avg	24.4	41.5	26.5	33.3	**18.42**	**21.81**	48.86	50.49	54.14	52.26
Std	22.4	26.5	15.8	**13.1**	**14.73**	14.39	17.60	16.83	28.41	31.65
Average (w/o P2)	11.5	26.4	17.3	29.2	**10.04**	**13.60**	40.45	43.94	57.98	54.67
Std (w/o P2)	02.3	04.7	**00.3**	12.7	02.85	**02.55**	11.41	14.35	31.89	36.22

Table 1. TREs (mm) for ADeLiR (`base`) and the baselines Opt [11] and LMR [10] for Manual and Automatic landmarks. * means w/o DA. Bold is best.

as Automatic deep Deformable Liver Registration (ADeLiR), which can be used with manual or automatical landmarks. We use ADeLiR (`base`) where `base` is the base architecture, namely ResNet or ViT (the encoder is ResNet34 or ViT-B-32 respectively, the base DA method is DANN) or TVT (the encoder is a modified ViT with additional attention modules [20] and the base DA method is TVT). We compare ADeLiR against two baselines: method `Opt` is an optimisation-based method [11], which is the current best for manual landmarks, and method `LMR` is a learning-based method [10], which is the current best for automatic landmarks.

We observe in table 1 that ADeLiR (ResNet) improves the performance of liver registration with manual landmarks by 6 mm or 24% on average. However, the strongest performance improvement is obtained for automatic landmarks with an average improvement of 11 mm and 13 mm, with and without Patient 2, representing 34% and 48%. The transformer encoders in ADeLiR (ViT) and ADeLiR (TVT) underperform; they do not capture the detailed image information, which is probably because they are patch-based. We thus continue the ablation study on ADeLiR (ResNet), shown in table 2. We observe that DA improves the accuracy by 18% for automatic landmarks and RTO improves the

accuracy by 9% for manual landmarks, with smaller impact for automatic landmarks. We observe that ADeLiR without DA outperforms ADeLiR on manual landmarks, indicating that DA does a good job of specialising the features to the automatic landmark domain, for which ADeLiR clearly outperforms. We precisely investigate the RTO sensitivity to landmark quality. We start from the clean landmarks to which we add noise with increasing level and measure the impact on the performance benefit of RTO on the tumour TRE. The noise is added by displacing the landmark points by a random white noise with controlled standard deviation in pixels. We found that RTO's efficiency starts decreasing beyond a 2 pixels noise level and tends to 0 beyond an 18 pixels noise level. This shows that RTO is stable, with reasonable sensitivity, contributing to improve the registration at the expected noise level of 7.6 pixels. The runtime for

Method	w/o DA, w/o RTO		w/o DA		w/o RTO		ADeLiR	
Segmentation	M	A	**M**	A	M	A	M	**A**
Patient 1	11.14	14.64	**06.48**	14.66	10.78	**12.17**	09.62	12.45
Patient 2	43.50	50.80	43.57	49.50	**42.36**	47.64	42.90	**46.44**
Patient 3	15.07	22.39	**13.46**	21.37	27.05	17.27	26.55	**17.13**
Patient 4	10.59	20.85	**10.18**	21.36	27.98	11.40	28.15	**11.20**
Avg	20.08	26.99	**18.42**	26.72	27.04	22.12	26.81	**21.81**
Std	13.63	13.64	14.73	**13.43**	**11.18**	14.91	11.80	14.39
Avg w/o P2	12.27	19.29	**10.04**	19.13	21.94	13.61	21.44	**13.60**
Std w/o P2	**02.00**	03.35	02.85	03.16	7.90	02.61	08.34	**02.55**

Table 2. Ablation study for ADeLiR (ResNet) using TREs (mm) with Manual and Automatic landmarks. Bold is best.

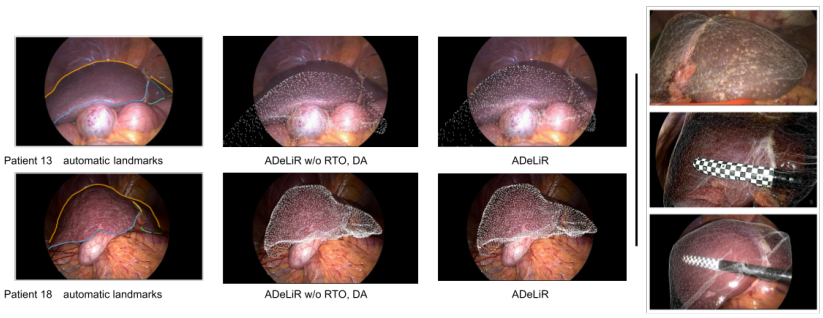

Fig. 4. Left: automatic segmentation and registration of the liver using ADeLiR with and without ablations. Right: extra ADeLiR results.

ADeLiR was 4.90 seconds (all images of both datasets; coarse registration time of 0.05 seconds, fine registration time of 4.85 seconds, with 10 average iterations

in RTO, on an Nvidia RTX 2080 Ti GPU without code optimisation). A 5 seconds registration is competitive with existing methods and compatible with the clinical usage.

6 Discussion and Conclusion

We have proposed an automatic liver registration method, which deforms a preoperative 3D liver model to fit an intraoperative 2D image. The method computes a coarse 3D affine transformation and refines it via a neural deformation model. We validate the method with a standard benchmark with ground truth. The proposed method, whilst using noisy automatic landmarks, reaches a registration accuracy on par with the state-of-the-art manual methods and largely outperforms the best existing automatic method. The proposed method has a reasonable computation time and is thus expected to form a valuable basis to develop concrete clinical use-cases of augmented reality in liver surgery. In future work, we have planned to increase the validation datasets, to avoid patient-specific training using knowledge-transfer from multiple patient models, and to improve landmark segmentation by training from a larger dataset.

Disclosure of Interests. Author E. Gadoux have no competing interests to declare that are relevant to the content of this article. Author A. Bartoli is the Chief Scientific Officer at SURGAR.

References

1. Adagolodjo, Y., Trivisonne, R., Haouchine, N., Cotin, S., Courtecuisse, H.: Silhouette-based pose estimation for deformable organs application to surgical augmented reality. IROS (2017)
2. Ali, S., Espinel, Y., Jin, Y., Liu, P., Güttner, B., Zhang, X., Zhang, L., Dowrick, T., Clarkson, J.M., Xiao, S., Wu, Y., Yang, Y., Zhu, L., Sun, D., Li, L., Pfeiffer, M., Farid, S., Maiere-Hein, L., Buc, E., Bartoli, A.: An objective comparison of methods for augmented reality in laparoscopic liver resection by preoperative-to-intraoperative image fusion. Medical Image Analysis (2024)
3. Bodenstedt, S., Wagner, M., Müller-Stich, B.P., Weitz, J., Speidel, S.: Artificial intelligence-assisted surgery: Potential and challenges. Visceral Medicine (2020)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. ICLR (2021). https://doi.org/arXiv:2010.11929
5. Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., March, M., Lempitsky, V.: Domain-adversarial training of neural networks. Journal of Machine Learning Research (2016)
6. Guan, H., Liu, M.: Domain adaptation for medical image analysis: A survey. IEEE Trans Biomed Eng (2022)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. CVPR (2016). https://doi.org/arXiv:1512.03385

8. K. Osumi, T.Y., Fujiyoshi, H.: Domain adaptation using a gradient reversal layer with instance weighting. International Conference on Machine Vision Applications (2019). https://doi.org/10.23919/MVA.2019.8757975
9. Koo, B., Özgür, E., Le Roy, B., Buc, E., Bartoli, A.: Deformable registration of a preoperative 3d liver volume to a laparoscopy image using contour and shading cues. Medical Image Computing and Computer Assisted Intervention (2017)
10. Labrunie, M., Pizzaro, D., Tilmant, C., Bartoli, A.: Automatic 3d/2d deformable registration in minimally invasive liver resection using a mesh recovery network. MIDL (2023)
11. Labrunie, M., Ribeiro, M., Mourthadhoi, F., Tilmant, C., Le Roy, B., Buc, E., Bartoli, A.: Automatic preoperative 3d model deformable registration in laparoscopic liver resection. Information Processing in Computer-Assisted Interventions (2022)
12. Long, M., Cao, Z., Wang, J., Jordan, M.I.: Conditional adversarial domain adaptation. NeurIPS (2018)
13. Mhiri, I., Pizzaro, D., Tilmant, C., Bartoli, A.: Neural patient-specific 3d-2d registration in laparoscopic liver resection. International Journal of Computer Assisted Radiology and Surgery (2024)
14. Plantefève, R., Haouchine, N., Radoux, J., Cotin, S.: Automatic alignment of pre and intraoperative data using anatomical landmarks for augmented laparoscopic liver surgery. Biomedical Simulation (2014)
15. Rabbani, N., Calvet, L., Espinel, Y., Roy, B.L., Ribeiro, M., Buc, E., Bartoli, A.: A methodology and clinical dataset with ground-truth to evaluate registration accuracy quantitatively in computer-assisted laparoscopic liver resection. Computer Methods in Biomechanics and Biomedical Engineering: Imaging Visualization (2021)
16. Remelli, E., Lukoianov, A., Richter, S., Guillard, B., Bagautdinov, T., Baque, P., Fua, P.: Meshsdf: Differentiable iso-surface extraction. In: Advances in Neural Information Processing Systems. vol. 33, pp. 22468–22478 (2020)
17. Robu, M., Ramalhinho, J., Thompson, S., Gurusamy, K., Davidson, B., Hawkes, D., Stoyanov, D., Clarkson, M.: Global rigid registration of ct to video in laparoscopic liver surgery. International J. of Comput Assist Radiol Surgs (2018)
18. Simo, J.C., Taylor, R.L.: Quasi-incompressible finite elasticity in principal stretches. continuum basis and numerical algorithms. Computer Methods in Applied Mechanics and Engineering **85**(3), 273–310 (1991)
19. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. European Conference on Computer Vision (2016)
20. Yang, J., Liu, J., Xu, N., Huang, J.: Tvt: Transferable vision transformer for unsupervised domain adaptation. WACV (2023)
21. Ziche, L., Lenzig, B., Bieck, R., Neumuth, T., Schoenfelder, S.: Image-based deep learning of finite element simulations for fast surrogate biomechanical organ deformations. Current Directions in Biomedical Engineering (2022)