



Learning Expected Appearances for Intraoperative Registration During Neurosurgery

Nazim Haouchine¹(✉), Reuben Dorent¹, Parikshit Juvekar¹, Erickson Torio¹,
William M. Wells III^{1,2}, Tina Kapur¹, Alexandra J. Golby¹, and Sarah Frisken¹

¹ Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA
nhaouchine@bwh.harvard.edu

² Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract. We present a novel method for intraoperative patient-to-image registration by learning Expected Appearances. Our method uses preoperative imaging to synthesize patient-specific expected views through a surgical microscope for a predicted range of transformations. Our method estimates the camera pose by minimizing the dissimilarity between the intraoperative 2D view through the optical microscope and the synthesized expected texture. In contrast to conventional methods, our approach transfers the processing tasks to the preoperative stage, reducing thereby the impact of low-resolution, distorted, and noisy intraoperative images, that often degrade the registration accuracy. We applied our method in the context of neuronavigation during brain surgery. We evaluated our approach on synthetic data and on retrospective data from 6 clinical cases. Our method outperformed state-of-the-art methods and achieved accuracies that met current clinical standards.

Keywords: Intraoperative Registration · Image-guided Neurosurgery · Augmented Reality · Neural Image Analogy · 3D Pose Estimation

1 Introduction

We address the important problem of intraoperative patient-to-image registration in a new way by relying on preoperative data to synthesize plausible transformations and appearances that are *expected* to be found intraoperatively. In particular, we tackle intraoperative 3D/2D registration during neurosurgery, where preoperative MRI scans need to be registered with intraoperative surgical views of the brain surface to guide neurosurgeons towards achieving a maximal safe tumor resection [22]. Indeed, the extent of tumor removal is highly correlated with patients' chances of survival and complete resection must be balanced against the risk of causing new neurological deficits [5] making accurate intraoperative registration a critical component of neuronavigation.

Most existing techniques perform patient-to-image registration using intraoperative MRI [11], CBCT [19] or ultrasound [9, 17, 20]. For 3D-3D registration,

3D shape recovery of brain surfaces can be achieved using near-infrared cameras [15], phase-shift 3D shape measurement [10], pattern projections [17] or stereo-vision [8]. The 3D shape can subsequently be registered with the preoperative MRI using conventional point-to-point methods such as iterative closest point (ICP) or coherent point drift (CPD). Most of these methods rely on cortical vessels that bring salient information for such tasks. For instance, in [6], cortical vessels are first segmented using a deep neural network (DNN) and then used to constrain a 3D/2D non-rigid registration. The method uses physics-based modeling to resolve depth ambiguities. A manual rigid alignment is however required to initialize the optimization. Alternatively, cortical vessels have been used in [13] where sparse 3D points, manually traced along the vessels, are matched with vessels extracted from the preoperative scans. A model-based inverse minimization problem is solved by estimating the model's parameters from a set of pre-computed transformations. The idea of pre-computing data for registration was introduced by [26], who used an atlas of pre-computed 3D shapes of the brain surface for registration. In [7], a DNN is trained on a set of pre-generated preoperative to intraoperative transformations. The registration uses cortical vessels, segmented using another neural network, to find the best transformation from the pre-generated set.

The main limitation of existing intraoperative registration methods is that they rely heavily on processing intraoperative images to extract image features (eg., 3D surfaces, vessels centerlines, contours, or other landmarks) to drive registration, making them subject to noise and low-resolution images that can occur in the operating room [2, 25]. Outside of neurosurgery, the concept of pre-generating data for optimizing DNNs for intraoperative registration has been investigated for CT to x-ray registration in radiotherapy where x-ray images can be efficiently simulated from CTs as digital radiographic reconstructions [12, 27]. In more general applications, case-centered training of DNNs is gaining in popularity and demonstrates remarkable results [16].

Contribution: We propose a novel approach for patient-to-image registration that registers the intraoperative 2D view through the surgical microscope to preoperative MRI 3D images by learning *Expected Appearances*. As shown in Fig. 1, we formulate the problem as a camera pose estimation problem that finds the optimal 3D pose minimizing the dissimilarity between the intraoperative 2D image and its pre-generated Expected Appearance. A set of Expected Appearances are synthesized from the preoperative scan and for a set of poses covering the range of plausible 6 Degrees-of-Freedom (DoF) transformations. This set is used to train a patient-specific pose regressor network to obtain a model that is texture-invariant and is cross-modality to bridge the MRI and RGB camera modalities. Similar to other methods, our approach follows a monocular single-shot registration, eliminating cumbersome and tedious calibration of stereo cameras, the laser range finder, or optical trackers. In contrast to previous methods, our approach does not involve processing intraoperative images which have several advantages: it is less prone to intraoperative image acquisition noise; it does not require pose initialization; and is computationally fast thus supporting real-

time use. We present results on both synthetic and clinical data and show that our approach outperformed state-of-the-art methods.

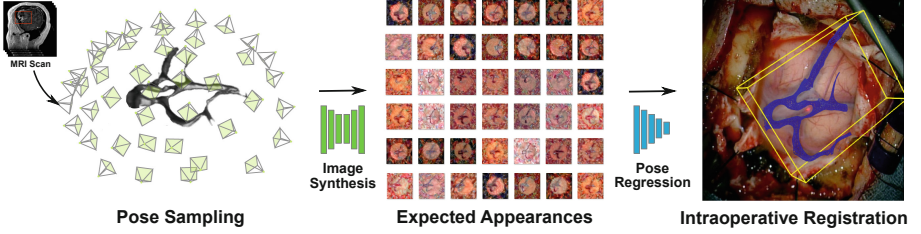


Fig. 1. Our approach estimates the 6-DoF camera pose that aligns a preoperative 3D mesh derived from MRI scans onto an intraoperative RGB image acquired from a surgical camera. We optimize a regressor network \mathcal{P}_Ω over a set of Expected Appearances that are generated by first sampling multiple poses and appearances from the 3D mesh using neural image analogy through \mathcal{S}_Θ .

2 Method

2.1 Problem Formulation

As illustrated in Fig. 1, given a 3D surface mesh of the cortical vessels \mathbf{M} , derived from a 3D preoperative scan, and a 2D monocular single-shot image of the brain surface \mathbf{I} , acquired intraoperatively by a surgical camera, we seek to estimate the 6-DoF transformation that aligns the mesh \mathbf{M} to the image \mathbf{I} . Assuming a set of 3D points $\mathbf{u} = \{u_j \in \mathbb{R}^3\} \subset \mathbf{M}$ and a set of 2D points in the image $\mathbf{v} = \{v_i \in \mathbb{R}^2\} \subset \mathbf{I}$, solving for this registration problem can be formalized as finding the 6-DoF camera pose that minimizes the reprojection error: $\sum_i^{n_c} \|\mathbf{A}[\mathbf{R}|\mathbf{t}] \begin{bmatrix} u_{c_i}^w \\ 1 \end{bmatrix} - v_i\|_2^2$, where $\mathbf{R} \in SO(3)$ and $\mathbf{t} \in \mathbb{R}^3$ represent a 3D rotation and 3D translation, respectively, and \mathbf{A} is the camera intrinsic matrix composed of the focal length and the principal points (center of the image) while $\{c_i\}_i$ is a correspondence map and is built so that if a 2D point v_i corresponds to a 3D point u_j where $c_i = j$ for each point of the two sets. Note that the set of 3D points \mathbf{u} is expressed in homogenous coordinates in the minimization of the reprojection error.

In practice, finding the correspondences set $\{c_i\}_i$ between \mathbf{u} and \mathbf{v} is non-trivial, in particular when dealing with heterogeneous preoperative and intraoperative modality pairs (MRI, RGB Cameras, ultrasound, etc.) which is often the case in surgical guidance. Existing methods often rely on feature descriptors [14], anatomical landmarks [13], or organ’s contours and segmentation [6, 18] involving tedious processing of the intraoperative image that is sensitive to the computational image noise. We alleviate these issues by directly minimizing the dissimilarity between the image \mathbf{I} and its Expected Appearance synthesized from \mathbf{M} .

By defining a synthesize function \mathcal{S}_Θ that synthesizes a new image $\hat{\mathbf{I}}$ given a projection of a 3D surface mesh for different camera poses, i.e. $\hat{\mathbf{I}} = \mathcal{S}_\Theta(\mathbf{A}[\mathbf{R}|\mathbf{t}], \mathbf{M})$, the optimization problem above can be rewritten as:

$$\operatorname{argmin}_{\mathbf{A}[\mathbf{R}|\mathbf{t}]} \left\{ \min_{\Theta} \|\mathbf{I} - \mathcal{S}_\Theta(\mathbf{A}[\mathbf{R}|\mathbf{t}], \mathbf{M})\| \right\} \quad (1)$$

This new formulation is correspondence-free, meaning that it alleviates the requirement of the explicit matching between \mathbf{u} and \mathbf{v} . This is one of the major strengths of our approach. It avoids the processing of \mathbf{I} at run-time, which is the main source of registration error. In addition, our method is patient-specific, centered around \mathbf{M} , since each model is trained specifically for a given patient. These two aspects allow us to transfer the computational cost from the intraoperative to the preoperative stage thereby optimizing intraoperative performance. The following describes how we build the function \mathcal{S}_Θ and how to solve Eq. 1.

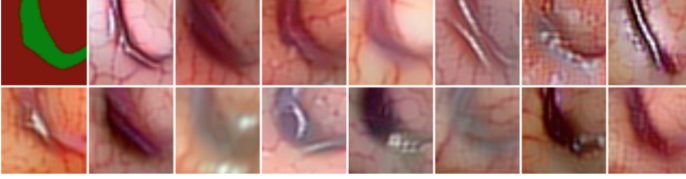


Fig. 2. An example of a set of Expected Appearances showing the cortical brain surface with the parenchyma and vessels. The network \mathcal{S}_Θ uses the binary image (top-left corner) computed from projecting \mathbf{M} using $[\mathbf{R}|\mathbf{t}]$ to semantically transfer 15 different textures $\{\mathbf{T}\}$ and synthesize the Expected Appearances $\{\hat{\mathbf{I}}\}$.

2.2 Expected Appearances Synthesis

We define a synthesis network $\mathcal{S}_\Theta : (\mathbf{A}[\mathbf{R}|\mathbf{t}], \mathbf{M}, \mathbf{T}) \rightarrow \hat{\mathbf{I}}$, that will generate a new image resembling a view of the brain surface from the 2D projection of the input mesh \mathbf{M} following $[\mathbf{R}|\mathbf{t}]$, and a texture \mathbf{T} . Several methods can be used to optimize Θ . However, they require a large set of annotated data [3, 24] or perform only on modalities with similar sensors [12, 27]. Generating RGB images from MRI scans is a challenging task because it requires bridging a significant difference in image modalities. We choose to use a neural image analogy method that combines the texture of a source image with a high-level content representation of a target image without the need for a large dataset [1]. This approach transfers the texture from \mathbf{T} to $\hat{\mathbf{I}}$ constrained by the projection of \mathbf{M} using $\mathbf{A}[\mathbf{R}|\mathbf{t}]$ by minimizing the following loss function:

$$\mathcal{L}_\Theta = \sum_l \sum_{ij} \left(\mathbf{w}_{\mathbf{T}_{\text{class}}}^{l,c} \mathcal{G}_{ij}^l(\mathbf{T}) - \mathbf{w}_{\mathbf{A}[\mathbf{R}|\mathbf{t}], \mathbf{M}}^{l,c} \mathcal{G}_{ij}^l(\hat{\mathbf{I}}) \right) \quad \text{for } c \in \{0, 1, 2\} \quad (2)$$

where $\mathcal{G}_{ij}^l(\mathcal{T})$ is the Gram matrix of texture \mathcal{T} at the l -th convolutional layer (pre-trained VGG-19 model), and $\mathbf{w}_{\mathbf{T}_{\text{class}}}^{l,c}$ are the normalization factors for each Gram matrix, normalized by the number of pixels in a label class c of $\mathbf{T}_{\text{class}}$. This allows for the quantification of the differences between the texture image \mathbf{T} and the generated image $\hat{\mathbf{I}}$ as it is being generated. Importantly, computing the inner-most sum over each label class c allows for texture comparison within each class, for instance: the background, the parenchyma, and the cortical vessels.

In practice, we assume constant camera parameters \mathbf{A} and first sample a set of binary images by randomly varying the location and orientation of a virtual camera $[\mathbf{R}|\mathbf{t}]$ w.r.t. to the 3D mesh \mathbf{M} before populating the binary images with the textures using \mathcal{S}_Θ (see Fig. 2). We restrict this sampling to the upper hemisphere of the 3D mesh to remain consistent with the plausible camera positions w.r.t. patient’s head during neurosurgery.

We use the L-BFGS optimizer and 5 convolutional layers of VGG-19 to generate each image following [1] to find the resulting parameters $\hat{\Theta}$. The training to synthesize for a single image typically takes around 50 iterations to converge.

2.3 Pose Regression Network

In order to solve Eq. 1, we assume a known focal length that can be obtained through pre-calibration. To obtain a compact representation of the rotation and since poses are restricted to the upper hemisphere of the 3D mesh (No Gimbal lock), the Euler-Rodrigues representation is used. Therefore, there are six parameters to be estimated: rotations r_x, r_y, r_z and translations t_x, t_y, t_z . We estimate our 6-DoF pose with a regression network $P_\Omega : \mathbf{I} \rightarrow \mathbf{p}$ and optimize its weights Ω to map each synthetic image \mathbf{I} to its corresponding camera pose $\mathbf{p} = [r_x, r_y, r_z, t_x, t_y, t_z]^\top$.

The network architecture of \mathcal{P}_Ω consists of 3 blocks each composed of two convolutional layers and one ReLU activation. To decrease the spatial dimension, an average pooling layer with a stride of 2 follows each block except the last one. At the end of the last hierarchy, we add three fully-connected layers with 128, 64, and 32 neurons and ReLU activation followed by one fully-connected with 6 neurons with a linear activation. We use the set of generated Expected Appearances $T^\mathcal{P} = \{(\mathbf{I}_i; \mathbf{p}_i)\}_i$; and optimize the following loss function over the parameters Ω of the network \mathcal{P}_Ω :

$$\mathcal{L}_\Omega = \sum_{i=1}^{|T^\mathcal{P}|} \left(\|\mathbf{t}_i - \hat{\mathbf{t}}_i\|_2 + \|\mathbf{R}_i^{\text{vec}} - \hat{\mathbf{R}}_i^{\text{vec}}\|_2 \right) \quad (3)$$

where \mathbf{t} and \mathbf{R}^{vec} are the translation and rotation vector, respectively. We experimentally noticed that optimizing these entities separately leads to better results. The model is trained for each case (patient) for 200 epochs using mini-batches of size 8 with Adam optimizer and a learning rate of 0.001 and decays exponentially to 0.0001 over the course of the optimization. Finally, at run-time, given an image \mathbf{I} we directly predict the corresponding 3D pose \mathbf{p} so that: $\mathbf{p} \leftarrow \mathcal{P}(\mathbf{I}; \hat{\Omega})$, where $\hat{\Omega}$ is the resulting parameters from the training.

3 Results

Dataset. We tested our method retrospectively on 6 clinical datasets from 6 patients (cases) (see Fig. 5). These consisted of preoperative T1 contrast MRI scans and intraoperative images of the brain surface after dura opening. Cortical vessels around the tumors were segmented and triangulated to generate 3D meshes using *3D Slicer*. We generated 100 poses for each 3D mesh (i.e.: each case) and used a total of 15 unique textures from human brain surfaces (different from our 6 clinical datasets) for synthesis using \mathcal{S}_θ . In order to account for potential intraoperative brain deformations [4] we augment the textured projection with elastic deformation [21] resulting in approximately 1500 images per case. The surgical images of the brain (left image of the stereoscopic camera) were acquired with a Carl Zeiss surgical microscope. The ground-truth poses were obtained by manually aligning the 3D meshes on their corresponding images.

We evaluated the pose regressor network on both synthetic and real data. The model training and validation were performed on the synthesized images while the model testing was performed on the real images. Because a conventional train/validation/test split would lead to texture contamination, we created our validation dataset so that at least one texture is excluded from the training set. On the other hand, the test set consisted of the real images of the brain surface acquired using the surgical camera and are never used in the training.

Metrics. We chose the average distance metric (ADD) as proposed in [23] for evaluation. Given a set of mesh's 3D vertices, the ADD computes the mean of the pairwise distance between the 3D model points transformed using the ground truth and estimated transformation. We also adjusted the default 5 cm-5 deg translation and rotation error to our neurosurgical application and set the new threshold to 3 mm-3 deg.

Accuracy-Threshold Curves. We calculated the number of 'correct' poses estimated by our model. We varied the distance threshold on the validation sets (excluding 2 textures) in order to reveal how the model performs w.r.t. that threshold. We plotted accuracy-threshold curves showing the percentage of pose accuracy variation with a threshold in a range of 0 mm to 20 mm. We can see in Fig. 3 that a 80.23% pose accuracy was reached within the 3 mm-3 deg threshold for all cases. This accuracy increases to 95.45% with a 5 mm-5 deg threshold.

Validation and Evaluation of Texture Invariance. We chose to follow a Leave-one-Texture-out cross-validation strategy to validate our model. This strategy seemed the most adequate to prevent over-fitting on the textures. We measured the ADD errors of our model for each case and report the results in

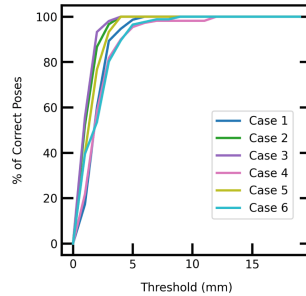


Fig. 3. Accuracy-threshold curves on the validation set.

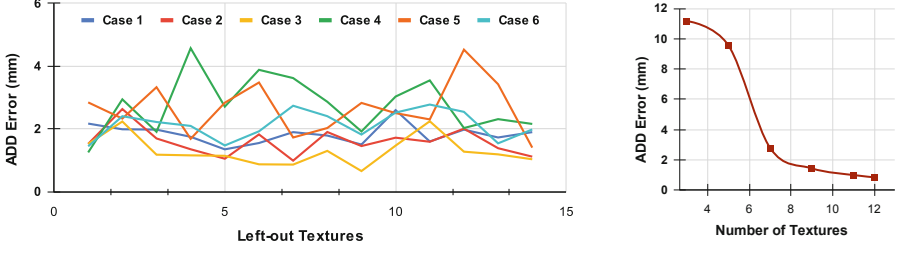


Fig. 4. Evaluation of texture invariance: (left) Leave-one-*Texture*-out cross validation and (right) impact of the number of textures on model accuracy.

Fig. 4-(left) for each left-out texture. We observed a variance in the ADD error that depends on which texture is left out. This supports the need for varying textures to improve the pose estimation. However, the errors remain low, with a 2.01 ± 0.58 mm average ADD error, over all cases. The average ADD error per case (over all left-out textures) is reported in Table 1. We measured the impact of the number of textures on the pose accuracy by progressively adding new textures to the training set, starting from 3 to 12 textures, while leaving 3 textures out for validation. We kept the size of the training set constant to not introduce size biases. Figure 4-(right) shows that increasing the number and variation of textures improved model performances.

Table 1. Validation on synthetic data and comparisons using real data.

Data	Experiment	Metric	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
<i>Synth.</i>	LOTO CV	Avg. ADD (mm)	1.80	1.56	1.27	2.83	2.58	2.07
	Acc./Thresh.	3mm-3deg (%)	89.33	96.66	98.07	81.81	93.33	80.23
<i>Real</i>	Ours	ADD (mm)	3.64	2.98	1.62	4.83	3.02	3.32
	ProbSEG	ADD (mm)	4.49	3.82	3.12	4.69	4.99	3.67
	BinSEG	ADD (mm)	9.2	4.87	12.12	8.09	11.43	6.29

Test and Comparison on Clinical Images. We compared our method (Ours) with segmentation-based methods (ProbSEG) and (BinSEG) [7]. These methods use learning-based models to extract binary images and probability maps of cortical vessels to drive the registration. We report in Table 1 the distances between the ground truth and estimated poses. Our method outperformed ProbSEG and BinSEG with an average ADM error of 3.26 ± 1.04 mm compared to 4.13 ± 0.70 mm and 8.67 ± 2.84 mm, respectively. Our errors remain below clinically measured neuronavigation errors reported in [4], in which a 5.26 ± 0.75 mm average initial registration error was measured in 15 craniotomy cases using intraoperative ultrasound. Our method outperformed ProbSEG in 5 cases out

of 6 and BinSEG in all cases and remained within the clinically measured errors *without* the need to segment cortical vessels or select landmarks from the intraoperative image. Our method also showed fast intraoperative computation times. It required an average of only 45 ms to predict the pose (tested on research code on a laptop with NVidia GeForce GTX 1070 8 GB without any specific optimization), suggesting a potential use for real-time temporal tracking.

Figure 5 shows our results as Augmented Reality views with bounding boxes and overlaid meshes. Our method produced visually consistent alignments for all 6 clinical cases without the need for initial registration. Because our current method does not account for brain-shift deformation, our method produced some misalignment errors. However, in all cases, our predictions are similar to the ground truth.

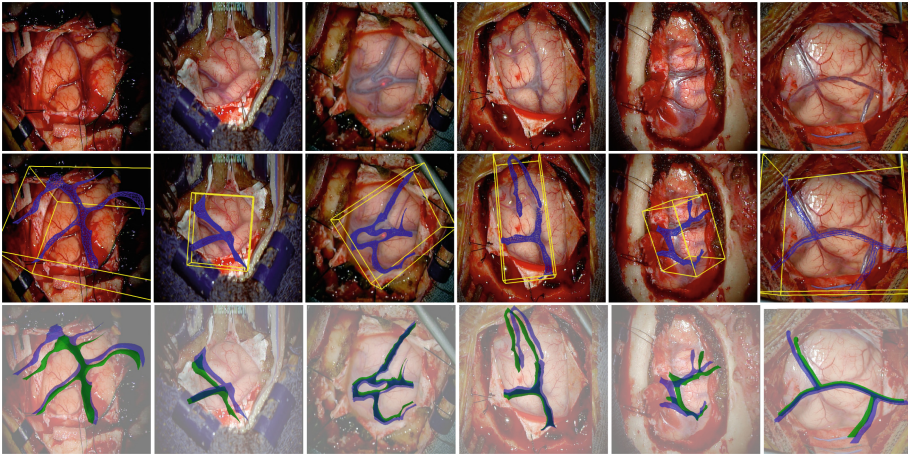


Fig. 5. Qualitative results on 6 patient datasets retrospectively showed in the first row. The second row shows Augmented Reality views with our predicted poses. The third row highlights 3D mesh-to-image projections using the ground-truth poses (green), and our predicted poses (blue). Our predictions are close to the ground truth for all cases. Note: microscope-magnified images with visible brain surface diameter ≈ 35 mm. (Color figure online)

4 Discussion and Conclusion

Clinical Feasibility. We have shown that our method is clinically viable. Our experiments using clinical data showed that our method provides accurate registration without manual intervention, that it is computationally efficient, and it is invariant to the visual appearance of the cortex. Our method does not require intraoperative 3D imaging such as intraoperative MRI or ultrasound,

which require expensive equipment and are disruptive during surgery. Training patient-specific models from preoperative imaging transfers computational tasks to the preoperative stage so that patient-to-image registration can be performed in near real-time from live images acquired from a surgical microscope.

Limitations. The method presented in this paper is limited to 6-DoF pose estimation and does not account for deformation of the brain due to changes in head position, fluid loss, or tumor resection and assumes a known focal length. In the future, we will expand our method to model non-rigid deformations of the 3D mesh and to accommodate expected changes in zoom and focal depth during surgery. We will also explore how texture variability can be controlled and adapted to the observed image to improve model accuracy.

Conclusion. We introduced Expected Appearances, a novel learning-based method for intraoperative patient-to-image registration that uses synthesized expected images of the operative field to register preoperative scans with intraoperative views through the surgical microscope. We demonstrated state-of-the-art, real-time performance on challenging neurosurgical images using our method. Our method could be used to improve accuracy in neuronavigation and in image-guided surgery in general.

Acknowledgement. The authors were partially supported by the following National Institutes of Health grants: R01EB027134, R03EB032050, R01EB032387, and R01EB034223.

References

1. Ulyanov, D., et al.: Texture networks: feed-forward synthesis of textures and stylized images. In: Proceedings of Machine Learning Research, vol. 48, pp. 1349–1357. PMLR (2016)
2. Maier-Hein, L., et al.: Surgical data science - from concepts toward clinical translation. *Med. Image Anal.* **76**, 102–306 (2022)
3. Fernandez, V., et al.: Can segmentation models be trained with fully synthetically generated data? In: Zhao, C., Svoboda, D., Wolterink, J.M., Escobar, M. (eds.) SASHIMI 2022. LNCS, vol. 13570, pp. 79–90. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16980-9_8
4. Frisken, S., et al.: A comparison of thin-plate spline deformation and finite element modeling to compensate for brain shift during tumor resection. *Int. J. Comput. Assist. Radiol. Surg.* **15**, 75–85 (2019). <https://doi.org/10.1007/s11548-019-02057-2>
5. González-Darder, J.M.: ‘State of the art’ of the craniotomy in the early twenty-first century and future development. In: González-Darder, J.M. (ed.) Trepanation, Trephining and Craniotomy, pp. 421–427. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22212-3_34
6. Haouchine, N., Juvekar, P., Wells III, W.M., Cotin, S., Golby, A., Frisken, S.: Deformation aware augmented reality for craniotomy using 3D/2D non-rigid registration of cortical vessels. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12264, pp. 735–744. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59719-1_71

7. Haouchine, N., Juvekar, P., Nercessian, M., Wells III, W.M., Golby, A., Frisken, S.: Pose estimation and non-rigid registration for augmented reality during neurosurgery. *IEEE Trans. Biomed. Eng.* **69**(4), 1310–1317 (2022)
8. Ji, S., Fan, X., Roberts, D.W., Hartov, A., Paulsen, K.D.: Cortical surface shift estimation using stereovision and optical flow motion tracking via projection image registration. *Med. Image Anal.* **18**(7), 1169–1183 (2014)
9. Ji, S., Wu, Z., Hartov, A., Roberts, D.W., Paulsen, K.D.: Mutual-information-based image to patient re-registration using intraoperative ultrasound in image-guided neurosurgery. *Med. Phys.* **35**(10), 4612–4624 (2008)
10. Jiang, J., et al.: Marker-less tracking of brain surface deformations by non-rigid registration integrating surface and vessel/sulci features. *Int. J. Comput. Assist. Radiol. Surg.* **11**, 1687–1701 (2016). <https://doi.org/10.1007/s11548-016-1358-7>
11. Kuhnt, D., Bauer, M.H.A., Nimskey, C.: Brain shift compensation and neurosurgical image fusion using intraoperative MRI: current status and future challenges. *Crit. Rev. Trade Biomed. Eng.* **40**(3), 175–185 (2012)
12. Lecomte, F., Dillenseger, J.L., Cotin, S.: CNN-based real-time 2D–3D deformable registration from a single X-ray projection. *CoRR abs/2003.08934* (2022)
13. Luo, M., Larson, P.S., Martin, A.J., Konrad, P.E., Miga, M.I.: An integrated multi-physics finite element modeling framework for deep brain stimulation: preliminary study on impact of brain shift on neuronal pathways. In: Shen, D., et al. (eds.) *MICCAI 2019. LNCS*, vol. 11768, pp. 682–690. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-32254-0_76
14. Machado, I., et al.: Non-rigid registration of 3d ultrasound for neurosurgery using automatic feature detection and matching. *Int. J. Comput. Assist. Radiol. Surg.* **13**, 1525–1538 (2018). <https://doi.org/10.1007/s11548-018-1786-7>
15. Marreiros, F.M.M., Rossitti, S., Wang, C., Smedby, Ö.: Non-rigid deformation pipeline for compensation of superficial brain shift. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) *MICCAI 2013. LNCS*, vol. 8150, pp. 141–148. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40763-5_18
16. Mildenhall, B., Srinivasan, P.P., Tancik, M., Barron, J.T., Ramamoorthi, R., Ng, R.: NeRF: representing scenes as neural radiance fields for view synthesis. *Commun. ACM* **65**(1), 99–106 (2021)
17. Mohammadi, A., Ahmadian, A., Azar, A.D., Sheykh, A.D., Amiri, F., Alirezaie, J.: Estimation of intraoperative brain shift by combination of stereovision and doppler ultrasound: phantom and animal model study. *Int. J. Comput. Assist. Radiol. Surg.* **10**(11), 1753–1764 (2015). <https://doi.org/10.1007/s11548-015-1216-z>
18. Nercessian, M., Haouchine, N., Juvekar, P., Frisken, S., Golby, A.: Deep cortical vessel segmentation driven by data augmentation with neural image analogy. In: *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 721–724. IEEE (2021)
19. Pereira, V.M., et al.: Volumetric measurements of brain shift using intraoperative cone-beam computed tomography: preliminary study. *Oper. Neurosurg.* **12**(1), 4–13 (2015)
20. Rivaz, H., Collins, D.L.: Deformable registration of preoperative MR, pre-resection ultrasound, and post-resection ultrasound images of neurosurgery. *Int. J. Comput. Assist. Radiol. Surg.* **10**(7), 1017–1028 (2015). <https://doi.org/10.1007/s11548-014-1099-4>
21. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) *MICCAI 2015. LNCS*, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

22. Sanai, N., Polley, M.Y., McDermott, M.W., Parsa, A.T., Berger, M.S.: An extent of resection threshold for newly diagnosed glioblastomas: clinical article. *J. Neurosurg.* **115**(1), 3–8 (2011)
23. Shotton, J., Glocker, B., Zach, C., Izadi, S., Criminisi, A., Fitzgibbon, A.W.: Scene coordinate regression forests for camera relocalization in RGB-D images. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2930–2937 (2013)
24. Skandarani, Y., Jodoin, P.M., Lalande, A.: GANs for medical image synthesis: an empirical study. *J. Imaging* **9**(3), 69 (2023)
25. Stoyanov, D.: Surgical vision. *Ann. Biomed. Eng.* **40**, 332–345 (2012). <https://doi.org/10.1007/s10439-011-0441-z>
26. Sun, K., Pheiffer, T., Simpson, A., Weis, J., Thompson, R., Miga, M.: Near real-time computer assisted surgery for brain shift correction using biomechanical models. *IEEE Transl. Eng. Health Med.* **2**, 1–13 (2014)
27. Tian, L., Lee, Y.Z., San José Estépar, R., Niethammer, M.: LiftReg: limited angle 2D/3D deformable registration. In: Wang, L., Dou, Q., Fletcher, P.T., Speidel, S., Li, S. (eds.) MICCAI 2022. LNCS, vol. 13436, pp. 207–216. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-16446-0_20