Unsupervised Method for Intra-patient Registration of Brain Magnetic Resonance Images based on Objective Function Weighting by Inverse Consistency: Contribution to the BraTS-Reg Challenge

Marek Wodzinski[1,2[0000−0002−8076−6246]], Artur Jurgas[1,2[0000−0002−9057−9428]],
Niccolò Marini[1,3[0000−0002−5273−5741]], Manfredo Atzori[1,4[0000−0001−5397−2063]],
and Henning Müller[1,5[0000−0001−6800−9878]]

[1] University of Applied Sciences Western Switzerland
Information Systems Institute, Sierre, Switzerland
[2] AGH University of Science and Technology
Department of Measurement and Electronics, Krakow, Poland
[3] Department of Computer Science, University of Geneva, Geneva, Switzerland
[4] Department of Neuroscience, University of Padova, Padova, Italy
[5] Medical Faculty, University of Geneva, Geneva, Switzerland
wodzinski@agh.edu.pl
marek.wodzinski@hevs.ch

Abstract. Registration of brain scans with pathologies is difficult, yet important research area. The importance of this task motivated researchers to organize the BraTS-Reg challenge, jointly with IEEE ISBI 2022 and MICCAI 2022 conferences. The organizers introduced the task of aligning pre-operative to follow-up magnetic resonance images of glioma. The main difficulties are connected with the missing data leading to large, nonrigid, and noninvertible deformations. In this work, we describe our contributions to both the editions of the BraTS-Reg challenge. The proposed method is based on combined deep learning and instance optimization approaches. First, the instance optimization enriches the state-of-the-art LapIRN method to improve the generalizability and fine-details preservation. Second, an additional objective function weighting is introduced, based on the inverse consistency. The proposed method is fully unsupervised and exhibits high registration quality and robustness. The quantitative results on the external validation set are: (i) IEEE ISBI 2022 edition: 1.85, and 0.86, (ii) MICCAI 2022 edition: 1.71, and 0.86, in terms of the mean of median absolute error and robustness respectively. Future work could transfer the inverse consistency-based weighting directly into the deep network training.

Keywords: Deep Learning · Image Registration · BraTS · BraTS-Reg · Brain Tumor · Glioma · Inverse Consistency · Missing Data.

1   Introduction

Registration of brain scans acquired using magnetic resonance imaging (MRI) is an active research area. Numerous works addressed the challenge by enforcing diffeomorphic deformations, using both classical [1], as well as deep learning-based (DL) approaches [4, 3, 8, 7]. However, it is still unclear what is the best approach for registering scans containing pathologies.

To answer this question, a Brain Tumor Sequence Registration (BraTS-Reg) challenge was organized in conjunction with IEEE ISBI 2022 and MICCAI 2022 conferences [2]. The challenge organizers proposed a task dedicated to registration of pre-operative MRI scans of patients diagnosed with a brain glioma, to follow-up scans of the same patient acquired after the treatment.

The surgical resection introduces the problem of missing data where the structure of interest is absent in one of the registered images. The problem of missing data in the registration of brain scans induces two main challenges: (i) large, nonrigid deformations caused by the tumor resection (depending on its size), and (ii) non-invertibility of the displacement field in the tumor area (the displacement field is unable to create new structures).

Contribution: In this work, we present our contribution to both the editions of the BraTS-Reg challenge [2]. We propose an unsupervised, hybrid approach, based on a deep neural network followed by an instance optimization (IO) attempting to address the challenge connected with large, nonrigid deformations. We introduce an additional, unsupervised inverse consistency-based weighting of the objective function. We present our motivations behind the method design and show the high quality of the registration in terms of the median absolute error (MAE) and robustness.

# 2   Methods

2.1   Overview and preprocessing

The proposed method consists of the following steps: (i) calculating the initial displacement field using a modified version of the LapIRN [9], (ii) tuning the displacement field by the instance optimization (IO) based, nonrigid registration method [11], and (iii - MICCAI edition only) an objective function weighting based on the inverse consistency. The ISBI 2022 pipeline is shown in Figure 1 and the MICCAI 2022 pipeline is shown in Figure 2.

The basic preprocessing was performed offline by challenge organizers and consisted of resampling the images to the same resolution. The processing pipeline starts with creating the source and target tensors. The tensors are created by concatenating all available modalities in the channel dimension, for both the source and target images. This results in two 4x240x240x155 volumes. The volumes are normalized channel-wise to [0-1] intensity values. We use the pre-operative scan as the source volume and the follow-up scan as the target volume. The challenge

design requires to warp the landmarks annotated in follow-up scans and we decided to not attempt to invert the noninvertible displacement fields, neither to voxelize the landmarks.
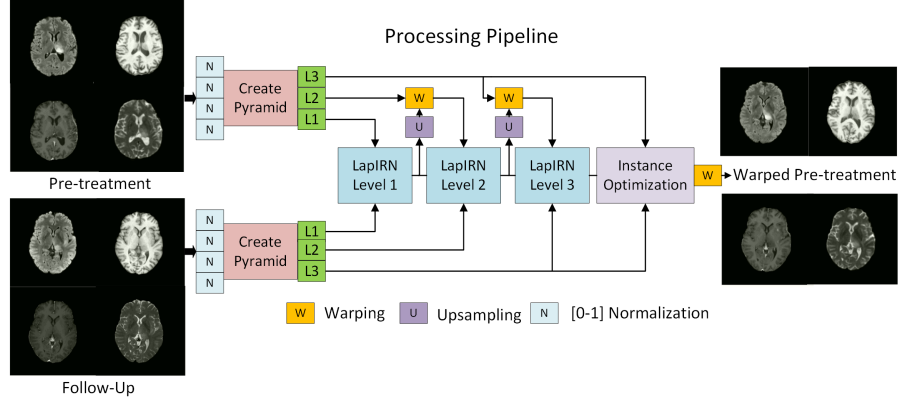
Fig. 1. The IEEE ISBI 2022 registration pipeline. Please note that some connections between LapIRN are omitted for the presentation clarity.
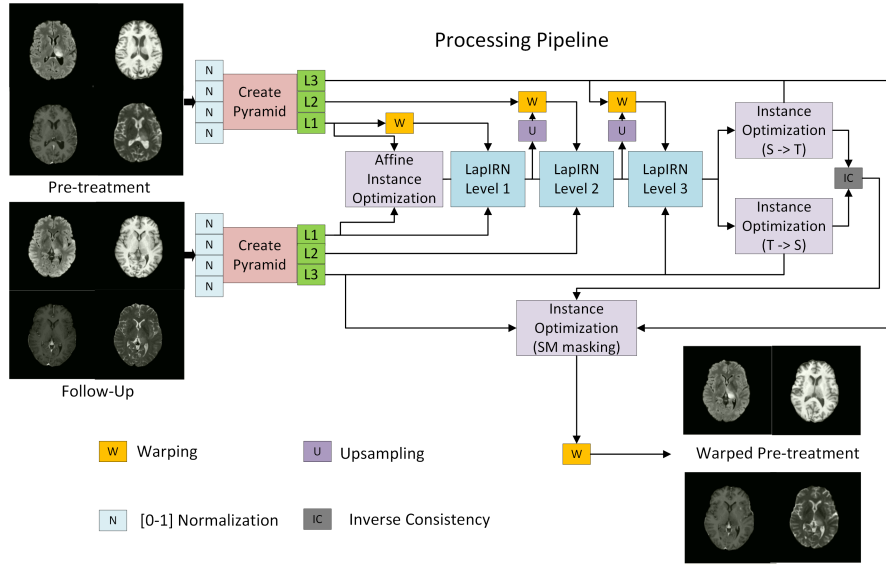
Fig. 2. The MICCAI 2022 registration pipeline. Please note that some connections between LapIRN are omitted for the presentation clarity. Note that during the 2nd edition of the challenge the pipeline starts with an initial affine registration.

2.2   Deep Network - LapIRN

The preprocessed volumes are directly passed to a slightly modified LapIRN [9]. The LapIRN is a multi-level network based on resolution pyramids to enlarge the receptive field. We decided to use the LapIRN because it won two editions of the Learn2Reg challenge by a huge margin and confirmed its ability to recover large, nonrigid deformations [6].

We use the LapIRN to optimize the following objective function:

$$O_{REG}(S, T, u) = \sum_{i=1}^{N} \frac{1}{2^{(N-i)}} (NCC(S_i \circ u_i, T_i) + \theta_i Reg(u_i)), \qquad (1)$$

where S_i, T_i are the pre-operative and follow-up scans at i-th resolution level respectively, u_i is the calculated displacement field, θ_i denotes the regularization coefficient at i-th level, NCC denotes the channel-wise normalized cross-correlation, Reg is the diffusive regularization, $\circ$ denotes the warping operation, and N is the number of pyramid levels.

We have reimplemented the network to have full control over the training process. We added group normalization due to the small batch size and deleted the scaling and squaring layers because we did not want to force diffeomorphic properties, nor to oversmooth the displacement fields. Even though it would be reasonable to calculate diffeomorphic displacements outside the tumor area, the smoothness of the deformation field was not taken into account to rank the submissions.

2.3   Instance Optimization

The LapIRN alone does not provide satisfactory results. Therefore, the displacement field calculated by LapIRN is passed as an initial transformation to a multi-level nonrigid instance optimization (IO) module. The module fine-tunes the displacement field by an iterative optimization implemented using PyTorch autograd. In the original formulation, the method optimizes the same cost function as the LapIRN.

We decided to use the IO because: (i) it improves the ability to register fine details since the displacement fields are optimized for each case separately, (ii) only the unsupervised objective function is used during LapIRN training, and (iii) the registration time is not scored during the challenge. The IO increases the registration time by several seconds which is usually acceptable in applications not requiring real-time alignment.

2.4   Inverse Consistency-based Weighting

To keep into account the missing data, we introduce an objective function weighting based on the inverse consistency. After the initial registration pipeline, the source and target images are bidirectionally registered again. The registration

is performed by several iterations of the IO. Then, the displacement fields are composed and the inverse consistency is calculated.

The inverse consistency is normalized to [0-1] range, raised to a given power, negated and smoothed using Gaussian filter with a predefined sigma. Both the sigma and the power are hyperparameters of the method. An exemplary visualization of the exemplary weighting mask is shown in Figure 3. Importantly, the weighting mask is applied both to the similarity measure and the regularization term. We decided to weight the regularization term because the displacement field is supposed to model tumor resection by the volume contraction which is prohibited by large values of the regularization coefficient.

Initially Warped
Pre-treatment   Follow-Up   Inverse Consistency
Error   Postprocessed
Weighting Function
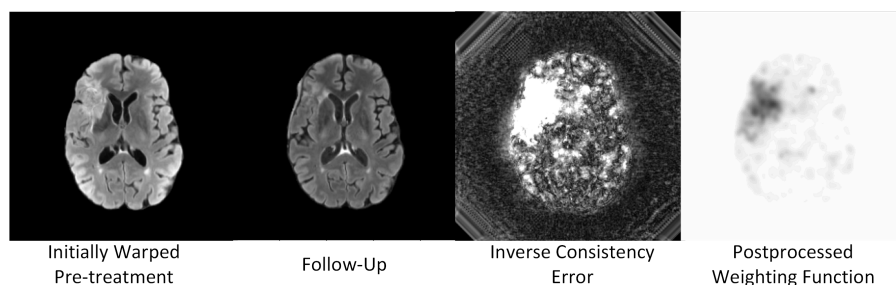
Fig. 3. Visualization of the objective function weighting. Note that small values of the weighting function are close to the tumor.

The intuition behind this approach is connected with the fact that regions without direct correspondences should have large values of the inverse consistency error. Thus, the information can be used to unsupervisedly detect non-corresponding regions and utilize this information in the objective function weighting.

2.5   Dataset and Experimental Setup

The multi-institutional dataset was curated and pre-processed retrospectively by the challenge organizers [2]. The dataset consists of pre-operative and follow-up pairs (of the same patient) diagnosed and treated for glioma. There are four MRI sequences for each pair: (i) T1-weighted (T1), (ii) contrast-enhanced T1-weighted (T1-CE), (iii) T2-weighted (T2), and (iv) fluid-attenuated inversion recovery (FLAIR).

The training set consists of 140 pairs annotated with landmarks varying from 6 to 50 per pair. The landmarks are defined using anatomical markers e.g. midline of the brain, blood vessel bifurcations, and anatomical shape of the cortex. We separated 5% of the training set as an internal validation set not used during the network training.

The external validation set consists of 20 pairs, following the structure of the training set, however, without releasing the landmarks for the pre-operative

scans. The evaluation on the external validation set is possible using the challenge infrastructure only. There is also an external test set, however, unavailable for the challenge participants. The test set results will be published in the follow-up challenge summary publication based on container submissions. Therefore, in this paper, we report results for the external validation set only.

All experiments are implemented in PyTorch, extended by PyTorch Lightning [5]. The LapIRN is trained using three pyramid levels. First, the first level is trained until convergence (with frozen weights of the remaining levels), followed by unfreezing the second level, and then finally the third level. We pre-train the network by registering the pairs in a cross-case manner, resulting in almost 18,000 training pairs. The network is then fine-tuned using the pairs representing scans of the same patient. Each iteration consists of registering 100 pairs with a batch size equal to 1. The weighted sum of local channel-wise NCC and diffusive regularization are used as the objective function. The pretraining is performed with a window size equal to 7 pixels, then changed to 3 pixels during the fine-tuning. The regularization coefficients are equal to 1000, 2000, and 4000 in the 1st, 2nd and 3rd pyramid level respectively. The initial learning rate is 0.003 and decays by a factor of 0.99 per iteration. No information about the anatomical landmarks is used during training.

The instance optimization, following the DL-based registration, is run using two resolution levels, for 40 and 20 iterations respectively. All the IO runs share the same hyperparameters. The objective function is the same as during training LapIRN and the hyperparameters are as follows: (i) NCC window size is equal to 3 pixels, (ii) the regularization coefficients are equal to 12500 and 25000 respectively, (iii) the initial learning rate is equal to 0.003. During the last instance optimization call, the inverse consistency is used to perform the similarity measure weighting. The sigma and the power of the inverse consistency map postprocessing are both equal to 2. The source code is openly available at [10].
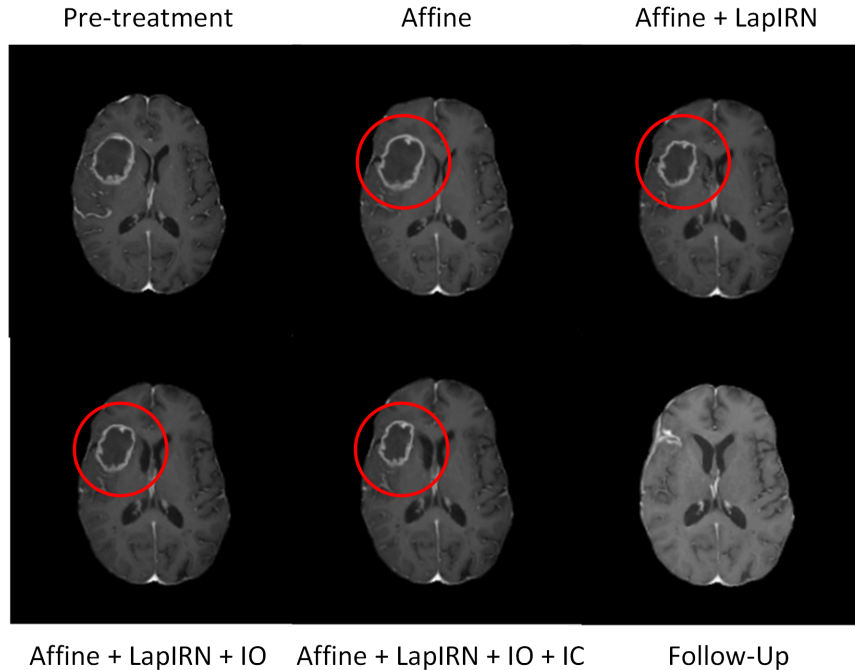
3   Results

The registration is evaluated quantitatively in terms of robustness and median absolute error (MAE) between the anatomical landmarks. The robustness is defined as the ratio of landmarks with improved alignment after registration to all landmarks. The challenge organizers decided to not score the registration time and the deformation field smoothness is used only in case of tie [2].

We show the MAE and robustness statistics on the external validation set in Table 1 and Table 2 for the IEEE ISBI and MICCAI editions respectively. We present the results before the registration, after the affine registration, for the IO and DL-based registration applied separately, for the proposed hybrid approach, and for the hybrid approach extended by unsupervised cost function weighting. These results come from the challenge evaluation platform. An exemplary visualization of the registration results are shown in Figure 4.

The proposed method scored the first place during the IEEE ISBI 2022 edition (however, without statistically significant differences compared to the re-

maining methods on the podium). Currently, the proposed method scores the second or third place on the external validation leaderboard for the MICCAI 2022 edition of the challenge depending on the evaluation run, without significant differences compared to the UZL team.

Fig. 4. Exemplary visualization of the registration results. Please note that the inverse consistency-based weighting improves the qualitative results close to the tumor volume. The tumor volume decreases and the quality of registration of fine-details around the tumor is improved. This improvement is not captured by the quantitative results based on the target registration error.

4   Discussion and Conclusion

The results confirm that the proposed hybrid approach, based on both deep learning and instance optimization (IO), improves the quantitative results compared to the methods used separately. Interestingly, the IO alone (extended by the IO-based affine registration) provides slightly better results than the LapIRN without any further tuning. This is expected because the IO optimizes each case separately and we do not use any supervision other than the unsupervised objective function to guide the training process. Interestingly, the worst robustness

Table 1. Quantitative results on the external validation set during the IEEE ISBI 2022 edition (MAE - Median Absolute Error).

Case	Initial		IO		LapIRN		LapIRN + IO	
ID	MAE	Robustness	MAE	Robustness	MAE	Robustness	MAE	Robustness
141	13.50	-	2.06	1.00	2.78	1.00	**1.94**	1.00
142	14.00	-	4.13	0.86	3.14	0.86	**2.82**	1.00
143	16.00	-	2.00	1.00	4.51	1.00	**1.76**	1.00
144	15.00	-	**3.54**	1.00	5.20	1.00	4.12	1.00
145	17.00	-	1.54	1.00	2.96	1.00	**1.19**	1.00
146	17.00	-	2.26	1.00	3.56	1.00	**2.24**	1.00
147	1.50	-	2.08	0.40	2.41	0.50	**1.79**	0.45
148	3.50	-	**1.30**	0.90	1.99	0.85	1.63	0.80
149	9.00	-	1.66	1.00	1.91	1.00	**1.62**	1.00
150	4.00	-	1.66	0.63	2.34	0.68	**1.32**	0.63
151	3.00	-	1.40	0.80	1.39	0.70	**1.17**	0.80
152	5.00	-	**1.43**	0.95	1.68	0.95	1.46	0.95
153	2.00	-	**1.47**	0.67	1.61	0.67	1.96	0.67
154	2.00	-	2.09	0.55	2.12	0.45	**2.03**	0.55
155	2.00	-	1.93	0.47	2.34	0.42	**1.92**	0.53
156	7.00	-	**1.74**	1.00	2.14	1.00	1.80	1.00
157	10.00	-	2.20	1.00	2.17	1.00	**1.22**	1.00
158	4.50	-	1.43	0.90	1.81	0.90	**1.19**	1.00
159	6.00	-	2.46	1.00	2.57	1.00	**2.18**	1.00
160	4.00	-	**1.59**	0.90	2.39	0.90	1.75	0.90
Mean	7.80	-	2.00	0.85	2.55	0.84	**1.85**	0.86
StdDev	5.62	-	0.72	0.20	0.95	0.20	**0.67**	0.19
Median	5.50	-	1.83	0.92	2.34	0.92	**1.77**	1.0
1st quartile	3.38	-	1.52	0.77	1.96	0.70	**1.42**	0.77
3rd quartile	13.63	-	2.12	1.00	2.82	1.00	**1.98**	1.00

is reported for cases with a low initial error, suggesting that the unsupervised registration cannot improve the results more.

The objective function weighting further improves the results. From the quantitative perspective the difference is not significant, however, it is connected with the characteristic of target registration error (TRE). The TRE is not dedicated to the evaluation of the image registration with missing data because the corresponding points cannot be annotated directly in the missing volume. Therefore, the correlation between the TRE and the qualitative results nearby the tumor is limited.

Our method has several limitations. First, we do not use information about sparse landmarks during training. We think that it would be possible to use the landmarks as a weak-supervision, however, with MAE at the level of 1.71 pixel, there is not much space for further improvements, and such a method would probably learn only the annotators' preferences. Second, we did not extensively explored the use of different modalities. Probably the use of modality used for

Table 2. Quantitative results on the external validation set during the MICCAI 2022 edition (MAE - Median Absolute Error). The initial errors are reported in Table 1.

Case ID	Affine MAE	Affine Robustness	LapIRN MAE	LapIRN Robustness	LapIRN + IO MAE	LapIRN + IO Robustness	LapIRN + IO + IC MAE	LapIRN + IO + IC Robustness
141	4.39	1.00	2.62	1.00	1.82	1.00	**1.82**	1.00
142	6.13	0.88	3.12	1.00	2.79	1.00	**1.92**	1.00
143	8.50	0.88	3.07	1.00	2.56	1.00	**1.88**	1.00
144	9.46	0.88	2.99	1.00	**2.31**	1.00	2.62	1.00
145	4.99	1.00	2.15	1.00	1.10	1.00	**1.08**	1.00
146	6.40	1.00	2.33	1.00	1.73	1.00	**1.66**	1.00
147	2.17	0.15	2.34	0.35	2.11	0.45	**1.64**	0.45
148	2.86	0.80	2.00	0.85	**1.57**	0.80	1.67	0.90
149	2.34	1.00	2.10	1.00	1.65	1.00	**1.61**	1.00
150	3.99	0.37	2.78	0.47	1.34	0.63	**1.28**	0.63
151	2.10	0.60	1.71	0.60	**1.30**	0.85	1.31	0.85
152	2.14	0.95	1.65	0.84	**1.43**	0.89	1.51	0.89
153	1.90	0.50	2.00	0.42	1.74	0.67	**1.66**	0.75
154	2.45	0.30	1.99	0.45	2.03	0.40	**1.98**	0.45
155	2.88	0.22	2.91	0.42	**2.01**	0.42	2.03	0.42
156	3.29	1.00	2.79	1.00	**1.48**	1.00	1.53	1.00
157	5.86	0.90	2.99	1.00	**1.52**	1.00	1.58	1.00
158	3.75	0.40	2.56	0.80	1.16	1.00	**1.16**	1.00
159	7.78	0.36	2.75	1.00	**2.30**	1.00	2.37	1.00
160	2.72	0.80	2.66	0.70	**1.79**	0.90	1.80	0.90
Mean	4.31	0.70	2.48	0.70	1.79	0.85	**1.71**	0.86
StdDev	2.32	0.30	0.46	0.25	0.46	0.22	**0.38**	0.21
Median	3.52	0.84	2.59	0.93	1.74	1.00	**1.66**	1.00
1st quartile	2.42	0.39	2.08	0.57	**1.47**	0.77	1.53	0.83
3rd quartile	5.93	0.96	2.82	1.00	2.05	1.00	**1.89**	1.00

the annotation would slightly improve the quantitative results. What is more, tuning the hyperparameters to just a subset of the available modalities could possibly further improve the results. Moreover, the inverse consistency weighting significantly increases the registration time (from several seconds to almost one minute). However, we think that future work could introduce the objective function weighting directly into the deep network training and significantly decrease the registration time.

In our opinion, several decisions regarding the challenge setup should be rethought. First, the MAE and the related robustness should not be used as the only evaluation criteria for image registration with missing data. The evaluation landmarks cannot be chosen directly within the tumor volume because it is missing in the target scan. Additional metrics, based e.g. on the relative tumor volume [11] could be used to evaluate the registration quality. Second, it could be considered to score the submissions using the displacement field smoothness outside the tumor volume (and the recurrences), as well as the registration time. Moreover, uploading the displacement fields directly instead of the transformed

landmarks would enable the organizers to calculate all the statistics automatically using the challenge evaluation platform.

To conclude, we proposed a hybrid approach based on the DL and IO as a contribution to the BraTS-Reg challenge. We have shown that the proposed method improves the alignment for majority of the external validation cases with high robustness. The proposed objective function weighting based on the inverse consistency further improved the results. The method could be further extended by directly incorporating the unsupervised objective function weighting during the deep network training.

References

1. Avants, B., et al.: Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. Medical Image Analysis **12**, 26–41 (2008)
2. Baheti, B., Waldmannstetter, D., Chakrabarty, S., et al.: The Brain Tumor Sequence Registration Challenge: Establishing Correspondence between Pre-Operative and Follow-up MRI scans of diffuse glioma patients. arXiv (2112.0697) (2021)
3. Balakrishnan, G., et al.: VoxelMorph: A Learning Framework for Deformable Medical Image Registration. IEEE Transactions on Medical Imaging **38**(8), 1788–1800 (2019)
4. Dalca, A., et al.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Medical Image Analysis **57**, 226–236 (2019)
5. Falcon, W., Cho, K.: A Framework For Contrastive Self-Supervised Learning And Designing A New Approach. arXiv preprint arXiv:2009.00104 (2020)
6. Hering, A., et al.: Learn2reg: comprehensive multi-task medical image registration challenge, dataset and evaluation in the era of deep learning. arXiv preprint arXiv:2112.04489 (2021)
7. Hoffmann, M., et al.: SynthMorph: Learning Contrast-Invariant Registration Without Acquired Images. IEEE Transactions on Medical Imaging (2021), Early Access
8. Mok, T., Chung, A.: Fast Symmetric Diffeomorphic Image Registration with Convolutional Neural Networks. IEEE CVPR pp. 4644–4653 (2020)
9. Mok, T., Chung, A.: Large Deformation Diffeomorphic Image Registration with Laplacian Pyramid Networks. MICCAI 2020 pp. 1–10 (2020)
10. Wodzinski, M., Jurgas, A.: Source Code. `https://github.com/MWod/BraTS_Reg` (2022)
11. Wodzinski, M., et al.: Semi-Supervised Deep Learning-based Image Registration Method with Volume Penalty for Real-time Breast Tumor Bed Localization. Sensors **21**(12), 1–14 (2021)