Response Letter

Dear Editors and Reviewers :

Thanks for your careful inspections and constructive comments on our submission titiled "A novel 3D Registration Network with Intra-class alignment for cross-domain Neonatal Brain MRI segmentation". We have revised the manuscript according to your comments. The reviewers comments and our point-to-point responses are listed as follows. All of the revised parts of our manuscript are in blue in the revised manuscript.

Response to Reviewer1(1QQm)

Summary: The authors propose a Transfer Layer (TL) to perform domain transfer following image registration to perform intensity transformation and shape registration in infant MRI, which exhibits large variation in shapes and intensity. The TL essentially changes the source domain image to match that of the target domain after a VoxelMorph-based registration network and then feeds this into a 3D U-net segmentation network. Results using infant MRI data from two datasets, NeoBrainS12 (n=7) and dHCP (n=40) demonstrate improved segmentation performance on 3D segmentation compared to other methods.

Strengths: The Transfer Layer (TL) appears to improve segmentation results compared to a joint-registration-segmentation approached without TL (Table 1) Comparison to two other unsupervised domain adaptation methods (MUNIT and CyCADA) provides good comparison to other methods.

Weaknesses: Methodological details are unclear in the text and do not provide sufficient description of the proposed Transfer Layer. The paper is challenging to read at times and could benefit from significant grammatical revisions.

Q1: *Sec. 3.2: The mean and variance images in the Transfer Layer calculation are not entirely clear to me. Are these created from the set of all training data and updated at each iteration? Otherwise, I am unsure where the mean value is coming from (and the variance).*

Response 1: Thanks for your suggestion. We have revised the explanation of the Transfer Layer and provided a detailed description for the mean and variance images in Line 168-174. The mean and variance images have the same size $L \times W \times H$ with the x_s or x_t. They can be calculated by equation 3 ($\overline{x}_s = \sum_{i=1}^{C} m_s^i \cdot y_s^i, \quad \widetilde{x}_s = \sum_{i=1}^{C} v_s^i \cdot y_s^i, \quad \overline{x}_t = \sum_{i=1}^{C} m_t^i \cdot y_s^i, \quad \widetilde{x}_t = \sum_{i=1}^{C} v_t^i \cdot y_s^i$). m_s^i,m_t^i,v_s^i and v_t^i are the i-th class mean and variance value with size $1 \times 1 \times 1$. y_s^i is the i-th class one-hot label of x_s with size $L \times W \times H$. And their specific calculation can be described as follows:

Line 160-165:

y_s^i is the i-th class one-hot label image of y_s with size $L \times W \times H$. m_s^i is the i-th class mean value of x_s and can be calculate as $m_s^i = \frac{\sum_{n=1}^{N} x_s(n) \cdot y_s^i}{\sum_{n=1}^{N} y_s^i}$. v_s^i is the i-th class variance value of x_s and can be calculate as $v_s^i = \sqrt{\frac{\sum_{n=1}^{N}(x_s(n) - \overline{x}_s(n))^2 \cdot y_s^i(n)}{\sum_{n=1}^{N} y_s^i(n)}}$. Similarly, for x_t and y_t, intra-class mean value can be calculated as $m_t^i = \frac{\sum_{n=1}^{N} x_t(n) \cdot y_t^i}{\sum_{n=1}^{N} y_t^i}$ and intra-class variance value can be calculated as $v_t^i = \sqrt{\frac{\sum_{n=1}^{N}(x_t(n) - \overline{x}_t(n))^2 \cdot y_t^i(n)}{\sum_{n=1}^{N} y_t^i(n)}}$.

Q2: *Sec. 3.2: Are the two MSE loss equations weighted equally? Is there a hyperparameter weighting these two loss functions?*

Response 2: Our improper expression causes this misunderstanding. The two MSE losses do not work at the same time, thus there is no weight assigned to them. The first MSE loss in equation 1 is used in step (i) to train the registration module without TL. Because the segmentation network is not trained at the beginning and cannot provide the pseudo-labels required by TL. The second MSE loss in equation 4 is used in step (iii) to train a complete registration module with TL. We revised this in Line 138-146 for a clear expression. And the step in which they are applied are stated before each equation.

Q3: *Sec. 2: A paper (Bo Li, et al. Ias-net: Joint intraclassly adaptive gan and segmentation network for unsupervised cross-domain in neonatal brain mri segmentation. Medical Physics, 2021) from the same authors is cited is similar. The main difference in this work appears to be the use of 3D networks compared to 2D previously. However, this is not clear in the current paper. For clarity, it would help reader to state the differences to this prior work from the same lab in the Background.*

Response 3:

Although this paper(RS-NET) and our previous work(IAS-NET) both perform the UDA in neonatal brain segmentation, they are quite different in many aspects:

(1) Different objectives between the two works. For the cross-domain neonatal brain segmentation task, RS-NET aims to make the UDA end-to-end, which can truly bridge the differences in both the shape and intensity across the domains. While IAS-NET is a UDA model, it cannot adapt to the large shape difference between domains. Thus, IAS-NET is not an end-to-end system and needs to perform complex preprocessing of registration to reduce the shape difference before UDA. This has been troubling us during the work of IAS-NET because it is difficult to apply to clinical practice.

(2) Different methods between the two works. IAS-NET is a GAN-based UDA model, which consists of a generator and multiple discriminators. Due to the semantic constraints, it can only perform intra-class intensity transfer but cannot complete shape transformation. While RS-NET is more concise. It consists of a registration network with TL layer, which simultaneously completes shape registration and intensity transfer.

(3) Different GPU cost between the two works. IAS-NET is a 2D model which integrates one generator and multiple discriminators and one segmentation network. Running on 2D images is already expensive, and it is almost impossible to expand to 3D images. But the cost of the proposed registration model in RS-NET is only 1/4 of that of the 3D segmentation model, which can be easily applied to the 3D segmentation network.

Regarding (1) and (3) mentioned above, we have revised the manuscript and illustrate the differences between the proposed RS-NET and the generative-based approaches(including ITA-NET) more clearly in Line 63-75. As for (2), we have revised the whole Method section for a better understanding about the proposed work.

Q4: *Sec. 5: Should a comparison to a standard U-net for image segmentation without registration be provided for an alternative baseline comparison segmentation approach? Otherwise, it is difficult to put the segmentation results in context since this is the goal of this paper.*

Response 4: We have added the segmentation result for image segmentation without registration in Table 1, in which the Baseline directly use original source-domain images and Scaling is a common scaling(scaling factor = 0.9) on them.

Q5: *Secs. 5 & 6: The GAN-based UDA model from previous work Li, et al 2021 is mentioned for comparison, but quantitative comparison results are not provided in this paper.*

Response 5: Our previous work IAS-NET is a 2D GAN-based UDA model for 2D segmentation network and the proposed 3D RS-NET is used for 3D segmentation network. Even the same UDA model would have large differences in the result between 3D and 2D segmentation network. Thus, it is unfair to quantitatively compare the two methods in Table 1. But, we supplemented the analysis of the differences in the performance between the proposed 3D work and the previous 2D work in our cross-domain segmentation task in Line 241-252.

Although the 3D model always achieve better results than the 2D model in the same-domain segmentation. Due to the scarce training samples, the shape diversity of 3D samples is largely limited. While the training samples of the 2D model are 48 times that of the 3D model, so this poor shape diversity is greatly relieved in the 2D case. Therefore, the shape difference between domains have much greater impact on the 3D model. The segmentation effect of 3D models in areas with large differences is not as good as that of 2D models. Additional training samples in the future can alleviate this problem of 3D cross-domain segmentation.

Q6: *Sec. 5: The Baseline registration method uses a scale transformation. Does this include translation, rotation, and scaling (an affine transformation)? Or, is it a isotropic scaling? Please clarify what is meant by scale transformation in this context. Some implementation details/citation would be helpful too.*

Response 6: Thanks for your advice. The expression of "scale transformation" is not correct. We have revised this sentence in Line 226-227.

Line 226-227:

The segmentation results are shown in Table 1, in which the 'Baseline' directly use original source-domain image and 'Scaling' is a common scaling(scaling factor = 0.9) on them.

Q7: *Table 1: It would be helpful and informative to report the standard deviation for these values to give readers a sense of variance in the results. (Same comment for Table 2 in the Appendix.)*

Response 7: We have repeated all experiments three times to calculate the standard deviation for the values in our experiments in Table 1 and Table 2.

Q8: *Figure 3: This figure is never referenced in the text. Are these images from two different subjects or from two different slices of the same subject? Also, a legend for the colormaps would be useful for readers.*

Response 8: Thanks for your advice. Figure 3 have been referenced in Line 238-239 and we have added description to each figure to illustrate these images in Figure 3 and Figure 4. For a better understanding of segmentation region, we have supplemented colormaps in Figure 3. Figure 4 contains more images and mainly shows the synthesized images, thus we do not add the colormaps in Figure 4.

Q9: *Sec. 5: How are the ground-truth brain segmentations created? Details would be helpful.*

Response 9: Thanks for your advice. The ground truth of NeobrainS12 datatset is provided directly. To obtain the ground truth of dHCP, a expert physician in our team make manual segmentations slice by slice for a week on a ROI Editor software. We have add this in Line 206-207.

Q10: *Grammar/typographical: Sec. 1, p2: wildly used → widely used Sec. 5, p8: slice thickness → I think you mean number of slices in each 3D volume here*

Response 10: Thanks for your advice. There are many Grammar errors and improper expressions in our manuscript. We have revised them carefully for every sentence in each paragraph and Figure.

Questions To Address In The Rebuttal:*Please see comments regarding TL methodology details above. In the manuscripts current format, I have a hard time understanding how the TL works according to equation 2. Also, details about how training utilizes the two MSE loss functions would be helpful.*

Response: Thanks for your advice. We have revised the whole paper and correct many Grammar errors carefully for a better understanding. Especially we rewrite the part of TL methodology.

Response to Reviewer2(DNgR)

Summary: The authors present a deep learning image segmentation approach to segment neonatal brain MRI images that jointly uses image registration. Segmentation and image registration of neonatal brain MRI images is a challenging task since the MRI intensity of the brains depends on the amount of myelination in the brain. The amount of myelination changes rapidly during the first two years of life. Thus, images of the same brain taken at different time may have very different MRI intensity profiles as well as shape differences.

Strengths: The novelty of this paper is the application to neonatal brain MRI images. This paper does a good job of explaining the challenges associated with segmenting and registering neonatal brain MRI images. The authors explain why 3D CNNs trained on one data set do not transfer to other clinical environments.

Weaknesses: This paper has poor grammar and is therefore difficult to read. There is little novelty in the method presented in this paper. The paper does cite the previous work Unsupervised Deformable Registration for Multi-Modal Images via Disentangled Representations by Chen Qin, Bibo Shi, Rui Liao, Tommaso Mansi, Daniel Rueckert, Ali Kamen which was presented at IPMI 2019. A preprint of this paper can be found at https://arxiv.org/abs/1903.09331 The method in Qin et al. is more advanced than what is presented in the current paper.

Questions To Address In The Rebuttal:*The authors need to cite the 2019 IPMI paper by Qin et al. and explain how their method is different or works better. The authors need to improve the readability of their paper by improving the grammar throughout the paper.*

Response: Thanks for your advice. I am sorry that due to our inappropriate writing the reviewer do not understand our novelty. There are many Grammar errors and improper expressions in our manuscript. We have revised them carefully for every sentence in each paragraph and

Figure. Especially we have rewritten the "Method" section and supplemented a lot of innovative content that we had previously omitted.

In addition, we have cited the work of Qin et al. (Qin et al., 2019) in Line 118-120. Although this paper(RS-NET) and current registration network both perform deep network to make registration, they are quite different in many aspects:

(1) Different objectives between them. The proposed RS-NET aims to make the unsupervised domain adaptation(UDA) in the cross-domain neonatal brain segmentation. Thus, we design that the RS-NET model can transfer both the shape and appearance across the domains. This is to make the segmentation network better adapt to the target domain in the case of great cross-domain difference. While current registration networks are to change the shape to the fixed image and maintain the appearance of moving image. For example, UMDIR(the work of Qin et al.) suggests a MUNIT-based method to decompose images into a shape representation and a appearance code for multi-modal image registration. It disentangles images into a shared shape space and different appearance spaces to align the moving image and maintain the invariant appearance without transferring.

(2) Different in methods between them. The proposed RS-NET designs a Transfer layer to make intra-class intensity transfer cross the domains at the same time as 3D registration. We have revised the explanation of the Transfer Layer and provided a detailed description for calculation of the mean and variance images in Line 168-175. This makes RS-NET accomplish the task of UDA for segmentation. While current registration networks focus on designing different model to transform image for a shape similarity to the fixed image. For example, UMDIR actually apply an image-to-image translation network for more accurate and fast alignment without a deformable field. But the proposed RS-NET is to train a designed network to output a deformable field for shape transformation and a TL for intensity transfer.

(3) Different GPU cost between them. Different from our previous 2D work(IAS-NET) (Li et al., 2021), the proposed RS-NET is designed to accomplish a end-to-end cross-domain 3D segmentation task without complex preprocessing of registration. This makes it easier to apply to the clinical practice. To ensure segmentation results, we implement a deeper network for segmentation and it is allocated most of the GPU resources. We design a small-structure registration model in RS-NET which is only 1/4 of that of the 3D segmentation model. Most of the current GAN-based models, if expanded to 3D ones, can not fulfill the end-to-end cross-domain 3D segmentation task.

Response to Reviewer3(5iXW)

Summary: The paper introduces a 3D image registration method for MRI images of infants. It deals with the challenge of registering T2-weighted MRI images that have been acquired with different scanners/settings. The network has two objectives: transforming the image and adapting the intensities (to the target domain). The idea of the paper is to use a combination of a registration and segmentation network. The objective of the segmentation network is to assign categorical labels to the data (white/gray matter etc.) so that it is easier to assign new intensities to the different tissue types. An experiment based on two different datasets used as source and target domain is performed showing superior performance. The paper addresses an important and difficult problem, but the technical presentation and the overall writing needs a major revision. It is difficult to understand and hard to evaluate in its current state.

Strengths: The paper addresses the difficult problem of cross-domain image registration and segmentation. I also highly appreciate the availability of source code. The abstract and first half of the introduction is well written.

Weaknesses: First, when reading the introduction and related work section, I had the impression the paper sells simultaneous segmentation and registration as a novelty. In my opinion, this is not true. Those two topics are strongly related, and I could find many related works such as A Cross-Stitch Architecture for Joint Registration and Segmentation in Adaptive Radiotherapy (MIDL 2020) or Training data-independent image registration using generative adversarial networks and domain adaptation (Pattern Recognition 2020).

I am not a native English speaker, and it is difficult for me to make detailed suggestions. But the writing needs to be improved. I found it difficult to understand the method in detail because of the writing style, the sudden deterioration of English, and the inconsistent names (R module, Register", register etc). Figure 1 - which is important - is confusing and difficult to understand. I feel that I cannot objectively and fairly evaluate their method as it stands.

Third, in my understanding, the algorithm needs categorial labels to work. Hits make sense for MRI images, where, for instance, Brain WM or GM needs to be aligned. I miss a discussion about the limitations of the algorithm in this regard.

In their experiments, they only applied a bias correction. In my opinion, the comparison is unfair then, since proper normalization of the inputs is crucial for the reference approaches (at least in terms of mean and standard deviation). It would also be nice to see a comparison with a "classical" registration approach like "ANTS" image registration with mutual information.

Q1: *First, when reading the introduction and related work section, I had the impression the paper sells simultaneous segmentation and registration as a novelty. In my opinion, this is not true. Those two topics are strongly related, and I could find many related works such as A Cross-Stitch Architecture for Joint Registration and Segmentation in Adaptive Radiotherapy (MIDL 2020) or Training data-independent image registration using generative adversarial networks and domain adaptation (Pattern Recognition 2020).*

Response 1: Thanks for your advice. Our improper expression causes this misunderstanding about our model, in which the segmentation and registration is not trained simultaneously. In fact, the two network are alternatively trained in our work. We have revised the whole "Introduction" and "Method" section for a better understanding for readers.

In addition, we have cited the works of Beljaards et al. (Beljaards et al., 2020)(MIDL 2020) and Mahapatra et al. (D.Mahapatra and Ge, 2020)(PR 2020) in Line 112-118. Although this paper(RS-NET) and current Joint Registration and Segmentation networks both perform deep network to make registration, they are quite different in many aspects:

(1) Different objectives. The proposed RS-NET aims to make the unsupervised domain adaptation(UDA) for cross-domain segmentation. Thus, we design the RS-NET to transfer both the shape and appearance across the domains for a better adaptation of the segmentation network to the target domain. While current joint registration and segmentation networks are to change the shape to the fixed image and maintain the appearance of moving image. For example, Beljaards et al. (Beljaards et al., 2020) aims to proposed a cross-stitch unit to exchange the information between a joint segmentation and registration network for better performance. Mahapatra et al. (D.Mahapatra and Ge, 2020) aims not to make segmentation but to make the

registration model adapt to another dataset. Thus, both of them cannot transfer the appearance of images to accomplish our UDA task for cross-domain segmentation.

(2) Different in methods. The proposed RS-NET designs a Transfer layer to make intra-class intensity transfer cross the domains at the same time as 3D registration, but the segmentation module is not trained simultaneously. We have revised the explanation of the Transfer Layer and provided a detailed description for calculation of the mean and variance images in Line 168-174. RS-NET is a generative-based model that the registration and segmentation models are trained alternatively. The segmentation model is trained by the images generated by the registration model for the adaptation to the target domain. While current Joint Registration and Segmentation networks are trained simultaneously to integrate information between the two networks for a better performance of both. They do not perform intensity transfer to make domain adaptation for segmentation.

(3) Different GPU cost. Different from our previous 2D work(IAS-NET) (Li et al., 2021), the proposed RS-NET is designed to accomplish a end-to-end cross-domain 3D segmentation task without complex preprocessing of registration. Thus it is easier to apply to the clinical practice. To ensure segmentation results, we design a small-structure registration model in RS-NET which is only 1/4 of that of the 3D segmentation model. Most of the current GAN-based models, if expanded to 3D ones, can not fulfill the end-to-end cross-domain 3D segmentation task.

Q2: *I am not a native English speaker, and it is difficult for me to make detailed suggestions. But the writing needs to be improved. I found it difficult to understand the method in detail because of the writing style, the sudden deterioration of English, and the inconsistent names (R module, Register", register etc). Figure 1 - which is important - is confusing and difficult to understand. I feel that I cannot objectively and fairly evaluate their method as it stands.*

Response 2: Thanks for your advice. There are many Grammar errors and improper expressions in our manuscript. We have revised them carefully for every sentence in each paragraph and Figure. In addition, the “Registration” in Figure 1 is the “R module” which denote the registration network with TL. And register represents the registration network without TL. We have revised “Registration” into “Registration module” to avoid this confusing.

Q3: *Third, in my understanding, the algorithm needs categorial labels to work. Hits make sense for MRI images, where, for instance, Brain WM or GM needs to be aligned. I miss a discussion about the limitations of the algorithm in this regard.*

Response 3: As mentioned above, our model with a small-structure registration network is to make UDA for segmentation. It is unreasonable to evaluate the region Hits and alignment of it. Thus, we analyzed its generated images in Figure 4 and the Mean Squared Error(MSE) and Structural Similarity(SSIM) is introduced to evaluate the similarity between the synthesized images and the target-domain images in Table 2.

Q4: *In their experiments, they only applied a bias correction. In my opinion, the comparison is unfair then, since proper normalization of the inputs is crucial for the reference approaches (at least in terms of mean and standard deviation). It would also be nice to see a comparison with a "classical" registration approach like "ANTS" image registration with mutual information.*

Response 4: Our improper expression causes this misunderstanding. The data preprocessing for all experiments is the same N4 correction and Sigmoid normalization. We have revised it in Line 134-135. In addition, our model is to make UDA for segmentation and it is designed for obtaining a intra-class similarity to the target-domain image. We mainly evaluate its segmentation performance and the similarity of the synthesized images to the target domain. Therefore, instead of comparing to the current registration methods, we are more interested in comparing with the state-of-art UDA methods.

Q5: *Equation (2): $TL(x_{st})$ looks like x_{st} is a parameter, and not the output.*

Response: We have revised this mistake in equation 2 and the text mentioned it in the manuscript.

Q6: *The paper claims that ... the above generative-based approaches also cannot work directly for the UDA task of neonatal brain segmentation. Firstly, the large shape differences between domains cannot be directly bridged by UDA methods. . Why the large shape differences can be addressed with the proposed approach but not with existing approaches? Existing registration approaches (like ANTS) are performing a multi-scale coarse-to-fine registration to deal with large displacements. The proposed method seems to not make use of any multi-scale analysis. Then how does it work so well?*

Response: Thanks for your advice. In fact, we have considered the impact of image scale differences between domains on cross-domain segmentation. But in our case, the segmentation model achieve limited gain from the scaling. Because the scale difference between domains is small. To illustrate this, we add a comparison to Table 1 in which 'Scaling' is the best scaling results(factor = 0.9) and does not improve much compared to the 'baseline'. Thus, the main factor affecting the UDA effect in our case is the difference in the shape and intensity distribution between domains, rather than scaling.

Q7: *Another claim is that ... registration network is always negatively affected by different intensity distribution between domains. Because the MSE loss tend to align the similar intensity to the same place, but these similar intensity voxels cross domains sometimes belong to different classes.. Then why not use mutual information?*

Response: This is an improper expression. Here we want to state is that the registration loss tends to align the voxels with similar intensity distribution to the same place regardless of their classes. We have revised this sentence in Line 70-71.

Q8: *It is also said that ..generative-based methods tend to synthesize the images similar to target as a whole but ignore the intra-class similarity. Can you elaborate more on this, please?*

Response: Thanks for your advice. "synthesize the images similar to target as a whole" is an improper expression. We have revised it in Line 71-73. The correct expression should be "Thirdly, current generative-based methods tend to achieve the intensity similarity to the target as a whole but ignore the intra-class one."

A novel 3D Registration Network with Intra-class alignment for cross-domain Neonatal Brain MRI segmentation

Bo Li[1] lb_whu@hust.edu.cn
Xinge You[1,3] youxg@mail.hust.edu.cn
Qinmu Peng[1,3] pengqinmu@hust.edu.cn
Jing Wang[2] jjwinflower@126.com
[1] *School of Electronic Information and Communication, Huazhong University of Science and Technology, Wuhan 430074, China*
[2] *Department of Radiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China*
[3] *Shenzhen Research Institute, Huazhong University of Science and Technology, Shenzhen 518000, China*

Editors: Under Review for MIDL 2022

Abstract

In neonatal brain Magnetic Resonance Image(MRI) segmentation, the model we trained on dataset from medical institutions fails to adapt to the clinical data. Because the clinical data(target domain) is largely different from the training dataset(source domain) in terms of scale, shape and intensity distribution. The registration network can transform the shape from the source domain to the target domain, while the intensity cannot be transferred. Current GAN-based unsupervised domain adaptation(UDA) models mainly focus on transferring the global intensity distribution, but cannot transform the shape and cover the intra-class similarity. In this case, we propose a joint Registration and Segmentation Network(RS-NET), both of which are trained to perform intra-class intensity and shape transformation and semantic segmentation. An adaptive transfer layer(TL) is designed to intra-classly transfer the intensity from source to target for reducing the cross-domain intensity difference, which can make the registration better align the multi-class regions. Meanwhile, the segmentation network can adapt to the target domain through the transformation of registration network. The experiment is carried out on two databases with big differences in shape, size and intensity distribution. The proposed method achieves state-of-the-art results in the compared UDA models for the 3D segmentation task. Source code (in TensorFlow) is available at: https://github.com/lb-whu/RS-NET/.

Keywords: Neonatal Brain Segmentation, 3D Registration, Unsupervised Domain Adaptation, MRI.

1. Introduction

The analysis of the neonatal brain morphological structure is crucial for assessing the brain development (Makropoulos et al., 2018a). Most of the brain disorders and diseases are reflected by the abnormalities of the morphological structure (Illavarason et al., 2018). An accurate and automatic segmentation of brain MRI images can help physicians better assess the neonatal brain and make therapeutic interventions. Among the current segmentation methods, the 3D convolutional neural network(CNN) achieves the best results (Xu et al., 2017; Nie et al., 2016).

However, 3D CNNs trained on specific database from medical institutions perform poorly in clinical practice. It is caused by the large cross-domain gap. This gap results mainly from the following three reasons: (1) The rapid development of the brain during the perinatal period leads to large shape difference between domains (Makropoulos et al., 2018a). (2) Different scanners, image acquisition protocols, and scanned subjects between the two domains result in the difference of intensity distribution (Tajbakhsh et al., 2020). (3) In the task of multi-class segmentation, the considerable cross-domain misalignments make the 3D segmentation worse. In addition, labeling all the tissues in the clinic practice is time-consuming. Therefore, we need to deal with an Unsupervised Domain Adaptation issue for the segmentation task.

Recently, many UDA methods have been used to make the cross-domain segmentation in some fields, such as Self-training, Adversarial Learning, and Generative-based Approach. Due to the large gap between the domains, most of the methods fail to apply to our case. For example, Self-training approach would produce the considerable misclassified pseudo-label which have negative effects on segmentation, even if with a suitable confidence threshold (Zhu et al., 2009). Adversarial-learning approach is weak in the case of the insufficient dataset because of its strategy of common feature refinement (Zhang et al., 2018). While, the dataset provided for training is always insufficient in brain MRI segmentation. Generative-based approach is more suitable to our case and it ,in fact, has been widely used in current UDA tasks for medical image segmentation (Oliver et al., 2018; Tajbakhsh et al., 2020).

CycleGAN is the most widespread one in the generative-based methods. Huo et al.(Huo et al., 2009) implement CycleGAN to synthesize CT distribution from labeled MR images for spleen segmentation in which an additional segmentation network is trained on the synthesized images for a better generalization. Chen et al. (Chen et al., 2019) combine the CycleGAN and segmentation network into a common encoder, in which it can receive the information from another modal(MRI) for cardiac CT segmentation. But, due to the large gap between domains, the cycle consistency of CycleGAN cannot completely preserve the semantic information of source images, which would lead to shape distortion (Zhang et al., 2018). Many works apply semantic constraint of segmentation network to force the generator of GAN to maintain the semantic information. For example, CyCADA (Hoffman et al., 2018) introduces FCN8s as a segmenter for providing semantic constraint for CycleGANs in cityscapes semantic segmentation. Zhang et al. (Zhang et al., 2018) apply Unet (Ronneberger et al., 2015) with a semantic consistency loss to CycleGAN for preserving semantic information of source domain.

However, the current generative-based approaches mentioned above cannot work directly for the UDA task of neonatal brain segmentation. Firstly, the large shape differences between domains cannot be directly bridged by the approaches. Therefore, the source-domain training images have to be registered before performing the UDA, which is time-consuming (Li et al., 2021). In fact, many current UDA works regard the registration as a preprocessing to overcome the shape difference between domains (Ackaouy et al., 2020; Li et al., 2021). Secondly, the registration network is always influenced by different intensity distribution between domains. Because the registration loss tends to align the voxels with similar intensity distribution to the same place regardless of their classes. Thirdly, current generative-based methods tend to achieve the intensity similarity to the target as a whole but ignore the intra-class one. Fourthly, the GPU memory consumption of current GAN-based models is too high, especially those 3D ones. Thus, they have to be trained independently apart from segmentation network which set barriers to clinical practice.

To address these issues, we propose a joint 3D Registration and Segmentation Network(RS-NET) for the cross-domain brain MRI segmentation. The Registration network with an Intra-class Transfer Layer(TL) is designed for transforming both the intensity and shape from the source domain to the target domain for the UDA task. In the training of Registration, the segmentation network provides pseudo-label for TL to calculate intra-class mean and variance of the target. Compared with other 3D GAN-based UDA models, the proposed registration network has much less GPU consumption, thus it can be trained simultaneously with the segmentation network and no longer be a complex preprocessing. As shown in the experimental result, the proposed method achieves the state-of-the-art results in the compared UDA models for the 3D segmentation task.

Our main contributions can be summarized as follows: (a) We designed a novel RS-NET for cross-domain neonatal brain MRI segmentation, which can transform both the intensity and the shape from source to target. The segmentation model can achieve better generalization in the target domain. (b) The proposed TL can achieve the intra-class similarity to the target domain and reduce the class misalignments of the registraion network. (c) The designed registration network only contains 1/4 parameters of the segmentation model, thus the whole framework(UDA and Segmentation) can be trained only once which is beneficial to clinical practice.

The rest of this paper is organized as follows. Firstly, we introduce the related work in Section II and our proposed work is presented in Section III. Secondly, experiment details and segmentation result are presented in Section IV and V. Thirdly, we summarize this paper in Section VI. Finally, the display of adaptive transfer result is supplemented in Appendix.

2. Related Work

Traditional registration methods align voxels between domains with enforced constraints by solving a pairwise optimization problem. This pairwise registration strategy requires intensive calculations, therefore its algorithms on CPU require hours to register only one pair of 3D images (Balakrishnan et al., 2019). But in most of the cases, deep learning networks for segmentation often require a large amount of training data, especially unlabeled target-domain data.

Recent registration methods apply the deep network on GPU to solve this issue. VoxelMorph (Balakrishnan et al., 2019) firstly propose a learning architecture, which uses a deep network to obtain the deformation field by training the network. Compared to the traditional pair-wised registration, this way is much time-saving. Marek Wodzinski et al. (Wodzinski et al., 2021) suggest a modified U-NET to make nonrigid image registration for real-time breast tumor bed localization, which is helpful to improve real-time radiation therapy after the tumor resection. Beljaards et al. (Beljaards et al., 2020) proposed a so-called cross-stitch unit to exchange the information between a joint segmentation and registration network for better performance of both in Adaptive Radiotherapy. Mahapatra et al. (D.Mahapatra and Ge, 2020) propose a GAN-based(generative adversarial networks) framework to ensure that the features extracted by the encoders are invariant to the input image type. Thus, their model trained on one dataset can give better registration performance for other datasets. Qin et al.(Qin et al., 2019) suggest a unsupervised multi-modal deformable image registration method(UMDIR) to align multi-modal images without deformable field by decomposing images into a shape representation and a appearance code. Compared to conventional registration approaches, this image-to-image translation based on MUNIT (Huang et al., 2018) make significant improvements in terms of both accuracy and speed. The current registration networks focus on designing different model to transform moving image for a shape similarity to the fixed image and maintain its appearance.

However, we wish to transfer both the shape and appearance for making the segmentation network better adapt to the target domain in the case of great cross-domain difference. Meanwhile, a small-structure model is expected to fulfill its UDA task in an end-to-end cross-domain 3D segmentation. Most of the current UDA models, if expanded to 3D to train with a 3D segmentation network, would exceed the memory of the GPU. Finally, the cross-domain intra-class intensity difference would cause many misalignments for the registration networks.

3. Method

The proposed RS-NET is composed of intra-class registration and segmentation network. The data preprocessing for all experiments is the same N4 correction and Sigmoid normalization. Experiments are conducted on the whole brain MR images without the Brain extraction(BET)(M.Smith, 2002).

3.1. Framework of Proposed RS-NET

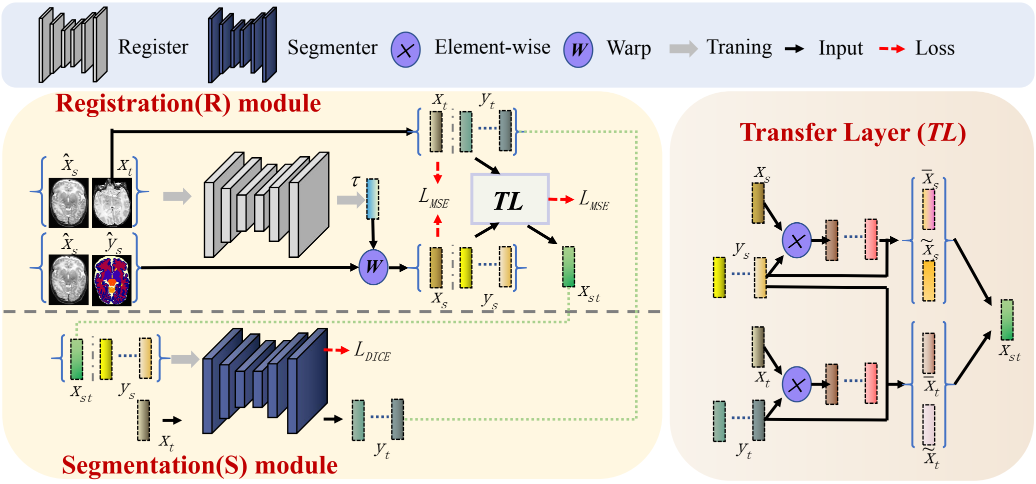

Figure 1: The overview of the proposed RS-NET.

As shown in Figure 1, both the registration network(R module) and the segmentation network(S module) are 3D U-NET structure. They are trained as follows: (i) R module without TL is firstly pre-trained by the source images $\hat{x}_s$ and target images x_t. (ii) The trained R module without TL transforms $\hat{x}_s$ and $\hat{y}_s$ to x_s and y_s. Then S module can be trained by the registered x_s and their corresponding label y_s. (iii) The R module with TL is initialized and retrained by $\hat{x}_s$ and x_t. Here, the S module trained in step (ii) can provide the pseudo-labels y_t of x_t for TL calculation. (iv) The R module with TL transforms the $\hat{x}_s$ and $\hat{y}_s$ into the registered image x_{st} and y_{st}. Finally, S module can be retrained by the UDA result x_{st} and y_{st}.

3.2. Registration Network with Intra-class Transformation

The R module aims to spatially transform source-domain images $\hat{x}_s$ with size $L \times W \times H$ into the ones whose shape and intensity distribution are consistent with the target-domain image x_t. But in step (i), R module without TL can only transform the shape. The volumetric input is a pair of $\hat{x}_s$ and x_t, in which they represents the moving and fixed image respectively. As mentioned above, in step (i), the TL is inactive and the R module is trained by MSE loss:

$$\mathcal{L}_{MSE} = \frac{\sum_{n=1}^{N}(x_t(n) - \tau(\hat{x}_s(n)))^2}{N} \qquad (1)$$

where $N = L \times W \times H$ is the voxel number of image, $\tau(\hat{x}_s(n)) = x_s$ is the transformed image.

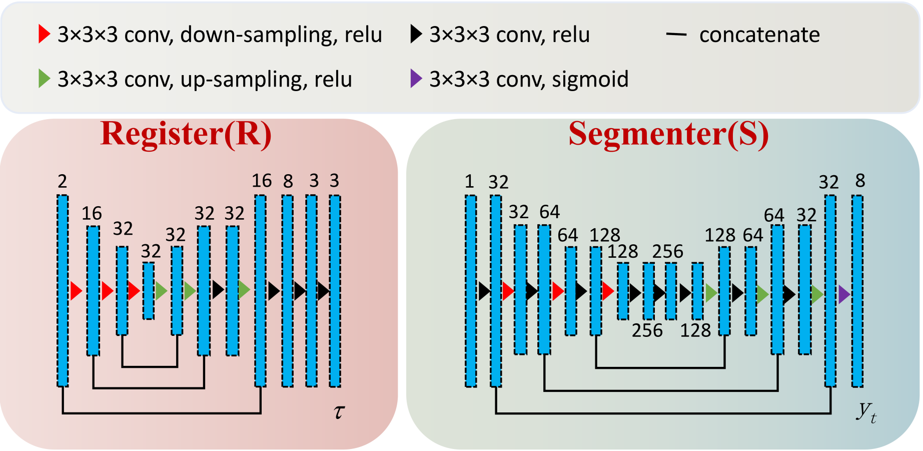

Figure 2: Structure of The Register and Segmenter.

In step (i), R module is trained to provide a coarse registration result for the training of segmentation in step (ii). The large difference of intra-class intensity and shape between the domains would lead to considerable misalignments. Because MSE loss tends to align the voxels with similar intensity regardless of their classes. Compared with directly training on $\hat{x}_s$, the S module trained on x_s can provide better pseudo-labels y_t for x_t which are used for TL calculation. Admittedly, the pseudo-labels are too inaccurate to be a criterion for segmentation. But it is sufficiently accurate to calculate the intra-calss mean and variance in step (iii). In step(iii), we introduce the TL to R module and initialize its parameter for retraining. To obtain the intra-class intensity similarity to the target domain, as shown in Figure 1, Transfer Layer(TL) calculation is proposed as follow:

$$TL(x_s, y_s, x_t, y_t) = \widetilde{x}_t \left(\frac{x_s - \overline{x}_s}{\widetilde{x}_s} \right) + \overline{x}_t \tag{2}$$

where x_s is the registered source-domain image by R module, $\overline{x}_s$ and $\widetilde{x}_s$ are respectively the intra-class mean and variance images of x_s, $\overline{x}_t$ and $\widetilde{x}_t$ are respectively the intra-class mean and variance images of x_t. They have the same size $L \times W \times H$. As shown on the right of Figure 1, they can be obtained as follows:

$$\overline{x}_s = \sum_{i=1}^{C} m_s^i \cdot y_s^i, \quad \widetilde{x}_s = \sum_{i=1}^{C} v_s^i \cdot y_s^i, \quad \overline{x}_t = \sum_{i=1}^{C} m_t^i \cdot y_s^i, \quad \widetilde{x}_t = \sum_{i=1}^{C} v_t^i \cdot y_s^i \tag{3}$$

where y_s^i is the i-th class one-hot label image of y_s with size $L \times W \times H$. m_s^i is the i-th class mean value of x_s and can be calculate as $m_s^i = \frac{\sum_{n=1}^{N} x_s(n) \cdot y_s^i}{\sum_{n=1}^{N} y_s^i}$. v_s^i is the i-th class variance value of x_s and can be calculate as $v_s^i = \sqrt{\frac{\sum_{n=1}^{N} (x_s(n) - \overline{x}_s(n))^2 \cdot y_s^i(n)}{\sum_{n=1}^{N} y_s^i(n)}}$. Similarly, for x_t and y_t, intra-class mean value can be calculated as $m_t^i = \frac{\sum_{n=1}^{N} x_t(n) \cdot y_t^i}{\sum_{n=1}^{N} y_t^i}$ and intra-class variance value can be calculated as $v_t^i = \sqrt{\frac{\sum_{n=1}^{N} (x_t(n) - \overline{x}_t(n))^2 \cdot y_t^i(n)}{\sum_{n=1}^{N} y_t^i(n)}}$. And thus, for training step(iii) the MSE loss is as follow:

$$\mathcal{L}_{MSE} = \frac{\sum_{n=1}^{N} (x_t(n) - TL(\tau(\hat{x}_s(n))))^2}{N} \tag{4}$$

where $TL()$ is the TL calculation.

The network of the R module is shown in Figure 2. To guarantee a better segmentation result, we design a deeper segmentation network and reduce the number of filters of the R module. The encoder consists of four resolution levels and the 3D convolutions with a kernel size of 3 and stride of 2. The decoder is composed of the 3D convolutions, the upsampling and the skip connections with the features learned from encoding layers. The whole network starts with 16 filters in the first layer. After each layer, a ReLU layer is followed. At the final layer, the feature is mapped to a three-dimension displacement vector as the deformable matrix τ.

3.3. Segmentation Network

As mentioned above, we apply a 3D U-NET for segmentation. As shown in Figure 2, the network starts with 32 filters in the first layer and each layer of the encoder or decoder is composed of two 3×3×3 convolutions, one batch normalization, one ReLU layer and the downsampling or the upsampling. Skip connections from encoder to decoder are also implemented. At the final layer, the output is mapped to a feature vector with eight components as the predicted one-hot label.

In step (iii), The S module is inactive and acts as a segmenter to provide pseudo-labels for R module. In stage (ii) and (v), it is trained as a segmentation network on the output of the R module. Dice loss is implemented in both steps as follow:

$$\mathcal{L}_{DICE} = \frac{1}{C} \cdot \sum_{i=1}^{C} (1 - \frac{2 f_{seg}^{i}(x_{st}) \cdot y_{s}^{i} + \gamma}{f_{seg}^{i}(x_{st})^2 + (y_{s}^{i})^2 + \gamma}) \tag{5}$$

where C denote the class number, f_{seg}^{i} is the prediction of the segmenter for the i-th class, y_{s}^{i} is the registered one-hot label.

4. Experiments

The datasets selected for the two different domains in our experiment are NeobrainS12(Isgum et al., 2015) and dHCP(Makropoulos et al., 2018b), which exist large difference in the shape and intensity distribution between them. NeobrainS12 dataset selects 40 weeks infants for MRI scan on Philips SENSE head 3T scanner. There are 7 T2-weighted axial MR images in NeobrainS12, in which resolution = 0.35×0.35×2 mm^3, TR/TE = 6293/120 ms, image size = 512×512×50. dHCP dataset selects 37-44 weeks infants for MRI scan on Philips Achieva 3T scanner. 40 T2-weighted MR images are provided, resolution = 0.8×0.8×0.8 mm^3, TR/TE = 12000/156 ms, image size = 290×290×155. For a easily labeling, we down-sample the dHCP images to 290×290×50. Then a expert physician make manual segmentations slice by slice for a week on a ROI Editor software. Finally, the images of the both datasets are cropped to the same size 272×208×48. NeobrinS12 is selected as the source domain and the dHCP as the target domain in our experiment.

The experimental detail is as follow. In step (i) and (iii), the learning rate for training register is initially used as $1{\times}10^{-5}$ and is then decreased to $1{\times}10^{-6}$ in 200 epochs. In step (ii) and (iv), the learning rate for training segmentation starts with $1{\times}10^{-3}$ and decayed to $1{\times}10^{-5}$ in 300 epochs. Both networks are using TensorFlow and Adam optimization and Momentum is 0.5 with a batch size of 1. We implement a standard workstation with an Intel Xeon (E5-2682) CPU and a NVIDIA TITAN X GPU 12G. For the proposed method, each step of training takes about 2 to 4 hours, and the entire training process can be completed in half a day.

5. Segmentation Result

We compare our proposed 3D model with the state-of-art UDA methods: MUNIT(Huang et al., 2018) and CYCADA(Hoffman et al., 2018). Both of them are trained on 2D slices of the registered NeoBrainS12(source) and dHCP(target). The 2D slices synthesized by them are composed of 3D images and then are trained on a same 3D Unet as the proposed method for segmentation test. Human brain is segmented into seven different classes: cortical gray matter(CGM), basal ganglia and thalami(BGT), white matter(WM), brain stem(BS), cerebellum(CB), ventricles(VENT), and cerebrospinal fluid(CSF).

The segmentation results are shown in Table 1, in which the 'Baseline' directly use original source-domain images and 'Scaling' is a common scaling(scaling factor = 0.9) on them. We display the average Dice Score and Sensitivity over three experiments. It can

Table 1: Segmentation Result

Method	CGM	WM	BGT	BS	CB	VENT	CSF	AVG
	Dice score(%)							
Baseline	76.5±1.8	54.5±4.1	82.9±2.2	18.3±3.1	65.9±2.5	53.6±1.7	69.5±1.5	60.2±1.8
Scaling	78.9±1.2	60.4±5.2	85.6±1.8	20.3±2.4	72.9±2.2	52.5±3.1	72.7±1.4	63.4±2.2
VM	87.7±1.4	74.1±2.8	90.5±1.5	74.2±1.9	85.6±1.5	65.3±2.5	82.5±1.6	79.7±1.1
VM&Munit	84.2±1.5	61.6±3.8	86.2±1.4	75.6±1.4	75.4±1.6	60.1±2.8	78.3±1.3	74.5±0.7
VM&Cycada	88.2±1.2	73.2±2.5	90.5±1.2	74.9±2.6	86.8±1.4	63.3±1.8	83.7±1.3	80.1±0.9
Proposed	**89.4**±1.1	**81.1**±1.7	**92.1**±0.9	**82.2**±1.8	**89.6**±1.2	**71.1**±2.1	**86.5**±1.5	**84.6**±0.8
Methed	Sensitivity(%)							
Baseline	73.8±2.1	48.7±5.8	90.3±1.9	15.6±2.7	58.1±2.4	56.6±2.2	85.3±1.4	61.2±1.9
Scaling	78.5±1.5	47.9±4.7	90.6±1.6	17.5±2.8	68.8±2.7	54.3±3.5	78.1±1.2	62.2±1.8
VM	89.8±1.2	67.7±3.2	90.4±1.8	64.2±2.1	78.5±1.6	73.4±2.4	86.2±1.8	78.6±1.7
VM&Munit	85.3±1.6	53.9±3.5	88.5±1.6	81.9±1.5	73.4±1.9	65.1±2.1	84.3±1.3	76.1±1.1
VM&Cycada	**90.4**±0.9	63.5±3.2	91.1±1.5	71.7±2.3	86.6±1.6	75.6±1.6	85.3±1.5	80.6±1.2
Proposed	89.2±1.3	**75.4**±2.0	**93.5**±0.5	**82.0**±1.6	**86.7**±1.4	**81.9**±2.3	**88.7**±1.3	**85.3**±0.9

be seen that without registration the 3D segmentation network trained on the 'Baseline' images completely fails in the target domain. Meanwhile, MUNIT reduces the Dice score of the segmentation network, which is due to the inconsistency between the synthesized images and its labels. And CYCADA has limited improvements to the segmentation network. This demonstrate that although the semantic constraints in CYCADA can maintain semantic consistency, it also limits the transfer effect to the target domain. The proposed RS-NET increases the average dice score by 5%, which shows that it greatly improves the generalization of the segmentation network in the target domain. As shown in Figure 3, the proposed method significantly outperforms the compared ones 'BS', 'CB' and 'VENT' regions. The synthesized images are shown in Figure 4 in the appendix.

In fact, the shape difference between domains has greater impact on 3D segmentation network than 2D ones. Compared to the same-domain segmentation, cross-domain task exists considerable misclassifications between WM and CB, VENT, BS regions which have large gap in shape. These misclassifications are more serious in the 3D model. It is why the accuracy of WM is lower than that of CGM in our experiment. Meanwhile, since the slice number of one sample is 48, the training samples of the 3D network are 48 times less than those of the 2D network. Thus, the shape diversity of 3D samples is largely limited. Compared to our previous 2D UDA work (Li et al., 2021), although the 3D RS-NET achieves better segmentation results in CGM and CSF regions which have small shape difference between domains, the segmentation results of other regions are even worse than that of the 2D network. We hope that the additional training samples for 3D cross-domain segmentation in the future can alleviate this issue.

6. Discussion and Conclusion

This work aims to train a 3D segmentation model to adapt it to the target domain in neonatal brain MR images. Experimental results shows that the shape and intensity distribution difference have a considerable impact on 3D cross-domain neonatal brain MRI

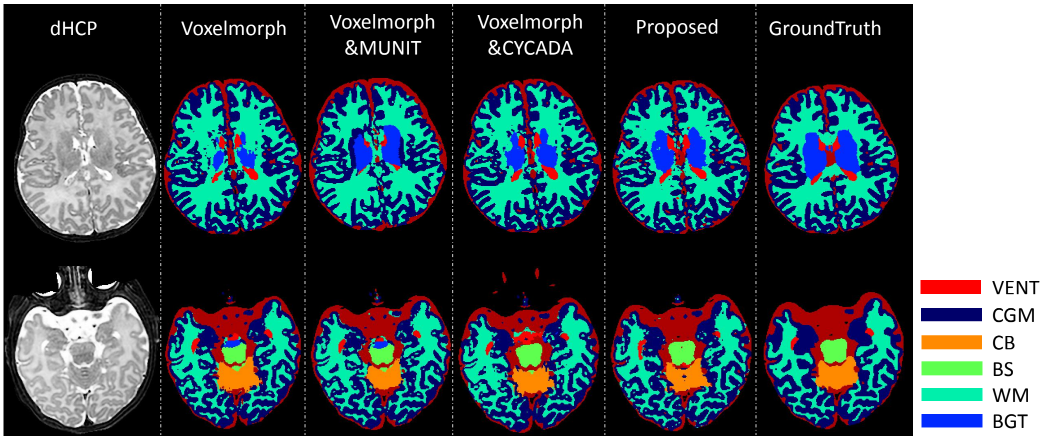

Figure 3: The display of Segmentation Results on two different slices from different subjects

segmentation. The proposed RS-NET can transform both the shape and intra-class appearance across the domains for a better adaptation of the segmentation network to the target domain. The intra-class intensity transformation by the proposed R module with TL can effectively increase the similarity of synthesized images to the target-domain images. Furthermore, most of the current UDA models, if expanded to 3D ones, can not fulfill the end-to-end cross-domain 3D segmentation task. Due to the small-structure design of R module, the proposed RS-NET can easily accomplish this task without complex preprocessing of registration, which is beneficial to clinical application.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant 11671161, Grant 61571205, and Grant 61772220, in part by the Key Program for International S&T Cooperation Projects of China under Grant 2016YFE0121200, in part by the Special Projects for Technology Innovation of Hubei Province under Grant 2018ACA135, in part by the Key Science and Technology Innovation Program of Hubei Province under Grant 2017AAA017, in part by the Natural Science Foundation of Hubei Province under Grant 2018CFB691, and in part by the Science, Technology and Innovation Commission of Shenzhen Municipality under Grant JCYJ20180305180637611,Grant JCYJ20180305180804836, and Grant JSGG20180507182030600.

References

A. Ackaouy, N. Courty, et al. Unsupervised domain adaptation with optimal transport in multi-site segmentation of multiple sclerosis lesions from mri data. *Frontiers in Computational Neuroscience*, 2020.

G. Balakrishnan, A. Zhao, et al. Voxelmorph: a learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38:1788–1800, 2019.

L. Beljaards, M. Elmahdy, F. Verbeek, and M. Staring. A cross-stitch architecture for joint registration and segmentation in adaptive radiotherapy. *Medical Imaging with Deep Learning*, 2020.

C. Chen, Q. Dou, et al. Synergistic image and feature adaptation: Towards cross-modality domain adaptation for medical image segmentation. *Association for the Advancement of Artificial Intelligence (AAAI)*, 2019.

D.Mahapatra and Zongyuan Ge. Training data independent image registration using generative adversarial networks and domain adaptation. *Pattern Recognition*, 2020.

Judy Hoffman, Eric Tzeng, et al. Cycada: Cycle-consistent adversarial domain adaptation. *International Conference on Machine Learning (ICML)*, pages 1994–2003, 2018.

Xun Huang, Ming-Yu Liu, et al. Multimodal unsupervised image-to-image translation. *IEEE European Conference on Computer Vision (ECCV)*, pages 172–189, 2018.

Y. Huo, Z. Xu, et al. Synseg-net: Synthetic segmentation without target modality ground truth. *IEEE Transactions on Medical Imaging*, 38:1016–1025, 2009.

P. Illavarason, J. R. Arokia, and P. K. Mohan. A study on the quality of life of cp children requiring early rehabilitation interventions. *International Conference on Science Technology Engineering and Mathematics(ICONSTEM)*, pages 91–97, 2018.

I. Isgum, Benders, et al. Evaluation of automatic neonatal brain segmentation algorithms: the neobrains12 challenge. *Medical Image Analysis*, 20:135–151, 2015.

Bo Li, Xinge You, Qinmu Peng, et al. Ias-net: Joint intraclassly adaptive gan and segmentation network for unsupervised cross-domain in neonatal brain mri segmentation. *Medical Physics*, 2021.

A. Makropoulos, S. J. Counsell, and D. Rueckert. A review on automatic fetal and neonatal brain mri segmentation. *NeuroImage*, 170:231–248, 2018a.

A. Makropoulos, Emma C.Robinson, et al. The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *NeuroImage*, 173:88–112, 2018b.

Stephen M.Smith. Fast robust automated brain extraction. *Human Brain Mapping*, 17(3): 143–155, 2002.

D. Nie, L. Wang, and Y. Gao. Fully convolutional networks for multi-modality isointense infant brain image segmentation. *International Symposium on Biomedical Imaging(ISBI)*, pages 1342–1345, 2016.

A. Oliver, A. Odena, et al. Realistic evaluation of deep semi-supervised learning algorithms. *arXiv preprint arXiv:1804.09170*, 2018.

C. Qin, Bibo Shi, R. Liao, Tommaso Mansi, D. Rueckert, and A. Kamen. Unsupervised deformable registration for multi-modal images via disentangled representations. *Information Processing in Medical Imaging*, 2019.

O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 234–241, 2015.

N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, et al. Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, 63, 2020.

M. Wodzinski, Izabela Ciepiela, et al. Semi-supervised deep learning-based image registration method with volume penalty for real-time breast tumor bed localization. *Sensors(Basel, Switzerland)*, 2021.

Y. Xu, T. Graud, and I. Bloch. From neonatal to adult brain mr image segmentation in a few seconds using 3d-like fully convolutional network and transfer learning. *IEEE International Conference on Image Processing(ICIP)*, pages 4417–4421, 2017.

Zizhao Zhang, Lin Yang, et al. Translating and segmenting multimodal medical volumes with cycle- and shape-consistency generative adversarial network. *IEEE conference on computer vision and pattern recognition (CVPR)*, pages 9242–9251, 2018.

X. Zhu, A. B. Goldberg, et al. Introduction to semi-supervised learning. *Synthesis lectures on artificial intelligence and machine learning*, pages 1–130, 2009.

Appendix A. Adaptive Transfer Result

The synthesized MR images by all compared methods and the corresponding input images of the two domains are shown in Figure 4. For a fair comparison of transferring, all the compared UDA methods use Voxelmorph as a preprocessing of registration. While the proposed method is to make end-to-end synthesis and segmentation. 'Voxelmorph Label' and 'Proposed Label' in Figure 4 refer to the transformed label image by Voxelmorph and the proposed method respectively. For better display, we enlarge a part of the synthesized images.

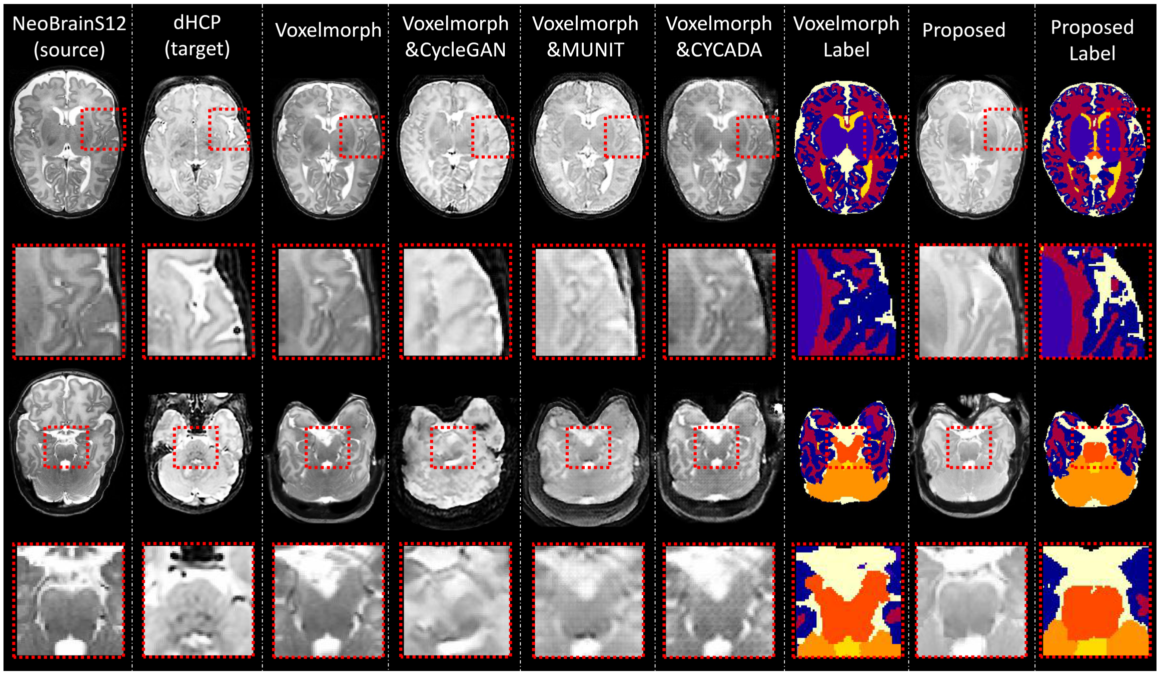

Figure 4: Synthesized T2 MR images by the compared methods. The images in the 1st and 3rd rows are from different subjects and the images in the same row are from the slice at the same position.

As shown in Figure 4, there are considerable shape and intensity difference between NeobrainS12 and dHCP. Thus GAN-based UDA model without a preprocessing of registration can not work in this case. As shown in the enlarged images of Figure 4, the images synthesized by the proposed method are more similar to the target domain than the compared methods. In addition, the synthesized images by proposed RS-NET shows the best appearance similarity to 'dHCP(target)' and maintains a good semantic consistency with 'Proposed Label'. Although the images synthesized by CYCADA well preserves the semantic content of the source domain, the appearance transfer is relatively poor. Unfortunately, due to the missing of supervision 'CycleGAN' and 'MUNIT' cannot maintain the semantic consistency with 'Voxelmorph Label'. Many shape distortions can be found in their synthesized images.

Table 2: The Adaptive Transfer Result

Indicator	**Voxelmorph**	**CycleGAN**	**CYCADA**	**MUNIT**	**Proposed**
MSE(target)	0.0154±0.0005	0.0152±0.0004	0.0245±0.0006	0.0144±0.0003	**0.0133**±0.0004
SSIM(target)	0.6039±0.0088	0.5812±0.0125	0.5534±0.0113	0.5877±0.0088	**0.6163**±0.0079
SNR	0.6812±0.0032	**0.7315**±0.0045	0.7063±0.0078	0.7176±0.0057	0.7218±0.0038

As shown in Table 2, we introduce Mean Squared Error(MSE) and Structural Similarity(SSIM) to quantitatively evaluate the effect of transferring to the target domain. We calculate the MSE distance and SSIM between all the synthesized images and their corresponding target-domain images. The average MSE value of each voxel and the average SSIM value of each sample are listed in Table 2. Table 2 shows that the images synthesized by the proposed method has the shortest MSE distance and the best structural similarity to the target domain. Finally, image SNR(Signal Noise Ratio) is used to measure the quality of the synthesized images. Since the SNR of the entire source-domain image is low, the SNR value of images registered by 'Voxelmorph' is much lower than other compared methods.