Unsupervised Domain Adaptation of MRI Skull-stripping Trained on Adult Data to Newborns

Abbas Omidi, Aida Mohammadshahi, Neha Gianchandani, Regan King, Lara Leijser, Roberto Souza
University of Calgary
Calgary, Alberta, Canada
{abbas.omidi, aida.mohammadshahi, neha.gianchandani
regan.king, lara.leijser, roberto.medeirosdeso}@ucalgary.ca

Abstract

Skull-stripping is an important first step when analyzing brain Magnetic Resonance Imaging (MRI) data. Deep learning-based supervised segmentation models, such as the U-net model, have shown promising results in automating this segmentation task. However, when it comes to newborn MRI data, there are no publicly available brain MRI datasets that come with manually annotated segmentation masks to be used as labels during the training of these models. Manual segmentation of brain MR images is time-consuming, labor-intensive, and requires expertise. Furthermore, using a segmentation model trained on adult brain MR images for segmenting newborn brain images is not effective due to a large domain shift between adult and newborn data. As a result, there is a need for more efficient and accurate skull-stripping methods for newborns' brain MRIs. In this paper, we present an unsupervised approach to adapt a U-net skull-stripping model trained on adult MRI to work effectively on newborns. Our results demonstrate the effectiveness of our novel unsupervised approach in enhancing segmentation accuracy. Our proposed method achieved an overall Dice coefficient of 0.916 ± 0.032 *(*$mean \pm std$*), and our ablation studies confirmed the effectiveness of our proposal. Remarkably, despite being unsupervised, our model's performance stands in close proximity to that of the current state-of-the-art supervised models against which we conducted our comparisons. These findings indicate the potential of this method as a valuable, easier, and faster tool for supporting healthcare professionals in the examination of MR images of newborn brains. All the codes are available at: https://github.com/abbasomidi77/DAUnet.*

1. Introduction

Brain Magnetic Resonance Imaging (MRI) is a non-invasive imaging technique that has become an essential tool for investigating the structure and function of the brain. Brain MRI plays a crucial role in the diagnosis and monitoring of various neurological disorders.

Skull-stripping is a crucial first step in neuroimaging analysis pipelines. Accurate segmentation can provide valuable information about the brain's structure, which is helpful in the diagnosis of neurological diseases. If a dataset with manual annotations, i.e., labels, is available, deep learning models can perform very well in automating this process, as they can learn to segment the brain from large amounts of labeled data. Automatic segmentation using deep learning models can save time and effort compared to manual segmentation, which is time-consuming and suffers from inter-/intra-subject annotation variability. It can also provide more consistent and reproducible results across different datasets and imaging protocols.

Automatically segmenting the brain in MR images can be particularly challenging for newborns. The neonatal brain is immature: The shape and size of the brain and its cortical and subcortical structures are different from adults. Also, image contrasts differ: On T1-weighted images, the intensity of the white matter (WM) is lower than that of the gray matter (GM). These image contrasts are the reverse of those seen in adults. This contrast difference is well known and caused by incomplete myelination in the white matter of the neonatal brain. Because of the variability of the myelination status in different fibers, the contrast between the gray and white matter in some areas is very poor [24]. Newborn MRI data are also more commonly affected by motion artifacts, which are another cause for distribution shifts between adult and newborn MRI data. Moreover, accurate segmentation of the neonatal brain from MR images is complex due to the lack of labeled newborn MRI segmentation datasets. These differences between adult and newborn MR images are illustrated in Fig. 1.

One possible approach to developing skull-stripping models for newborn brain MRI could involve training mod-

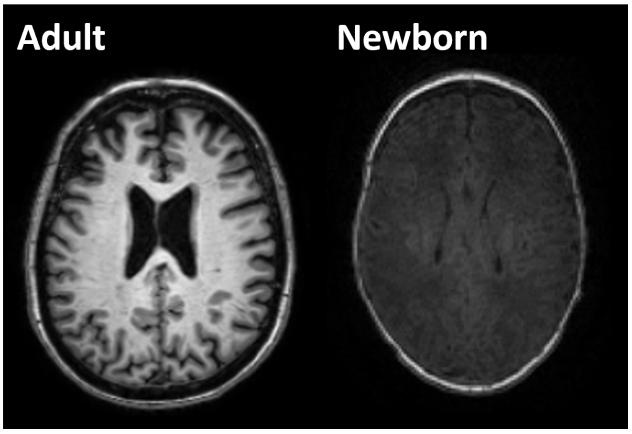


Figure 1: T1-weighted MR images of an adult subject (left) and newborn subject (right). Notice that the WM and GM tissue contrast is inverted and has poorer contrast on the newborn scans. The newborn image also has noticeable motion artifacts.

els on adult brain data since there are many publicly available labeled datasets [30, 22, 28], and then use these models to obtain brain segmentations on newborn MRI. However, a significant hurdle arises here from the fact that there is a large domain shift between adult and newborn brain MRI data. Therefore, these models solely trained on adult data do not generalize to newborns.

To overcome the challenge of performing skull-stripping of newborn MR images, we introduce a novel unsupervised domain adaptation (DA) technique capable of using labeled adult MRI data combined with unlabeled newborn MRI data to develop a skull-stripping method that segments newborn MR images accurately. Our proposed approach applies a special data augmentation transformation by inverting the adults' MRI contrast to reduce the domain shift between the adult and newborn data combined along with introducing a U-Net-based [26] domain adversarial training approach that makes our model learn domain invariant features [12]. To the best of our knowledge, this is the first time an unsupervised Domain-Adversarial Neural Network (DANN) method [12] has been investigated to address the challenges of accurately skull-stripping the neonatal brain from MR images.

Our novel unsupervised DA approach, comprising a combination of contrast inversion and a 3D U-Net-based adversarial DA method, is evaluated on both datasets of presumed healthy adult and newborn T1-weighted MR images, and the results show that this approach significantly improves the quality of skull-stripping in newborns' MR images, while preserving the quality of the segmentation in adult MR images. These results highlight the potential of combining deep learning models with DA techniques to address the challenges of domain shift in medical imaging and improve the accuracy of skull-stripping in newborn MR images.

2. Related work

Domain adaptation techniques play a pivotal role in enhancing the robustness and generalization of segmentation models across diverse datasets. These methods aim to address domain shift, where variations in image acquisition protocols, modalities, and resolutions hinder the seamless transfer of models from source to target domains [33, 20]. Adversarial domain adaptation has emerged as a powerful approach in which a domain discriminator is simultaneously trained with a segmentation network. This framework encourages the segmentation network to generate features that are domain-invariant, minimizing the distribution gap between domains [11]. Another avenue for tackling domain shift is domain-invariant feature learning, which focuses on extracting features that are insensitive to domain-specific variations. This allows models to generalize effectively across different datasets [6]. The integration of such domain adaptation techniques and feature learning methods into segmentation tasks, including tasks like skull-stripping, enhances the adaptability of models to variations in imaging domains and modalities, promoting more accurate and reliable results.

Skull-stripping refers to the process of segmenting the brain from non-brain tissue in MRI scans. To address the challenges of manual skull-stripping, various automated methods have been proposed in the literature. These methods can be broadly classified into two categories: traditional image-processing techniques and deep learning-based methods. Traditional image processing techniques include BEaST [10], region growing [21], and deformable models [7]. However, these methods are limited by their dependence on predefined features and parameters and can be affected by the variability in MR images.

Recently, deep learning-based methods have gained popularity for skull-stripping in MR images due to their ability to learn complex patterns from the data [25, 19]. Convolutional Neural Networks (CNNs) have shown promising results in various medical imaging tasks, especially in image segmentation [23]. For example, Brebisson *et al.* proposed a fully convolutional neural network (FCN) for skull-stripping in MR images and showed improved accuracy compared to traditional image processing techniques [8]. Chen *et al.* proposed a deep residual network for skull-stripping in MR images and showed that the proposed method outperforms traditional methods in terms of accuracy and efficiency [5].

However, one of the challenges in using deep learning-based methods for skull-stripping in MR images is the presence of domain shift between the training and testing data.

Domain shift refers to the difference in the distribution of the data between the source domain and the target domain. In the case of skull-stripping in MR images, the domain shift can be due to the difference in image acquisition protocols, image resolution, or image modalities. To address this challenge, recent studies have proposed the use of DA techniques to improve the generalization of deep learning models to new domains [35, 2, 27].

In the context of addressing the challenges associated with domain shift in contrast and resolution, Billot *et al.* proposed a novel technique utilizing a generative model [3]. This model was trained on a big dataset containing various extensive datasets existing in the literature [34, 17]. By leveraging this generative model, realistic synthetic brain MRI scans were generated, closely resembling the appearance of real scans. The synthetic scans were subsequently employed to train a segmentation model, and their results show efficacy in segmenting real brain MRI scans across a wide range of contrasts and resolutions.

In the area of newborns' skull-stripping, Weisenfeld *et al.* developed an automatic segmentation algorithm for brain MRI of newborn infants based on taking aligned T1- and T2-weighted MR images and assigning a label to each voxel in the image [36]. Research has shown that with intensity correction, edge-preserving noise smoothing, and tissue classification, good results in the skull-stripping of newborns can be achieved [29]. However, the authors acknowledge that it is difficult to distinguish cortical grey matter from myelin in the newborns' brain tissue. To the best of our knowledge, no research can be found in the literature regarding adapting segmentation models trained on adult data to newborns.

In comparing and contrasting the existing approaches within the realm of skull-stripping for brain MRIs, it becomes evident that addressing domain shift and the scarcity of annotated newborn data remains a critical challenge. Traditional methods, such as BEaST, region growing, and deformable models, have been fundamental but often lack the adaptability required to handle the inherent variability in MR images. Deep learning-based approaches, including convolutional neural networks (CNNs) like the FCN and deep residual networks, have revolutionized skull-stripping for adult brain MRIs, showcasing improved accuracy over traditional techniques. However, the domain shift that arises when transitioning to newborn brain images introduces a substantial obstacle. Our proposed unsupervised adaptation approach stands out as a promising solution to this issue. By integrating domain adaptation techniques inspired by recent advancements, our method offers a pathway to effectively adapt a U-net model, originally trained on adult MRI data, to accurately segment newborn brain images. Leveraging insights from generative models in the handling of domain shift further solidifies the efficacy of our approach. As a result, we contribute to addressing the unique challenges of newborn skull-stripping, ultimately improving the efficiency and precision of segmentation tasks for neonatal brain MRIs.

3. Materials and Methods

3.1. Dataset

3.1.1 Adult Dataset

In our study, we utilized the Calgary-Campinas public brain MRI dataset [30], which has 359 (176M:183F) T1-weighted, 3D, 1 mm isotropic adult brain MRIs. The dataset also provides skull-stripped, i.e., brain masks, and WM, GM segmentation masks. These data were collected in MRI scanners from different vendors (General Electrica [GE], Philips, Siemens) at two different magnetic fields (1.5 T and 3 T).

3.1.2 Newborn Dataset

We used a private dataset collected at the Alberta Children's Hospital using a GE 3T MRI scanner. The dataset has 12 (7F:5M) T1-weighted, 3D, 1 mm $\times$ 0.5 mm $\times$ 0.5 mm newborn brain MRIs. Brain masks were manually obtained on ten sagittal slices spaced by ten slices each on 5 out of the 12 samples in the dataset for evaluation purposes.

It is worth noting that the manual labeling process of an entire MRI volume is slow. It requires more than 8 hours for one MRI volume. This highlights the significance of developing an automated model for skull-stripping newborn brains in MRI scans that does not require labeled data for training.

3.2. Proposed Method

3.2.1 Overview

The visual overview in Fig. 2 illustrates the architecture of our network, which includes two learning loops to improve the segmentation accuracy of newborns' brains in MR images. The first loop is a standard 3D U-net learning loop that operates on the adult MRI data (source domain images) and their corresponding labels. In the forward propagation, the network creates the brain masks for adults' brains, and in the backpropagation process, the network parameters are updated to refine the segmentation accuracy.

The second loop focuses on the discriminator, which is added after the bottleneck. The main objective of the discriminator is to distinguish between the source domain images and the newborn MRI (target domain images). During the backpropagation process, the discriminator is updated properly to improve its ability to differentiate between the two types of images. However, when the gradients reach the encoder, a gradient reversal layer is activated, which

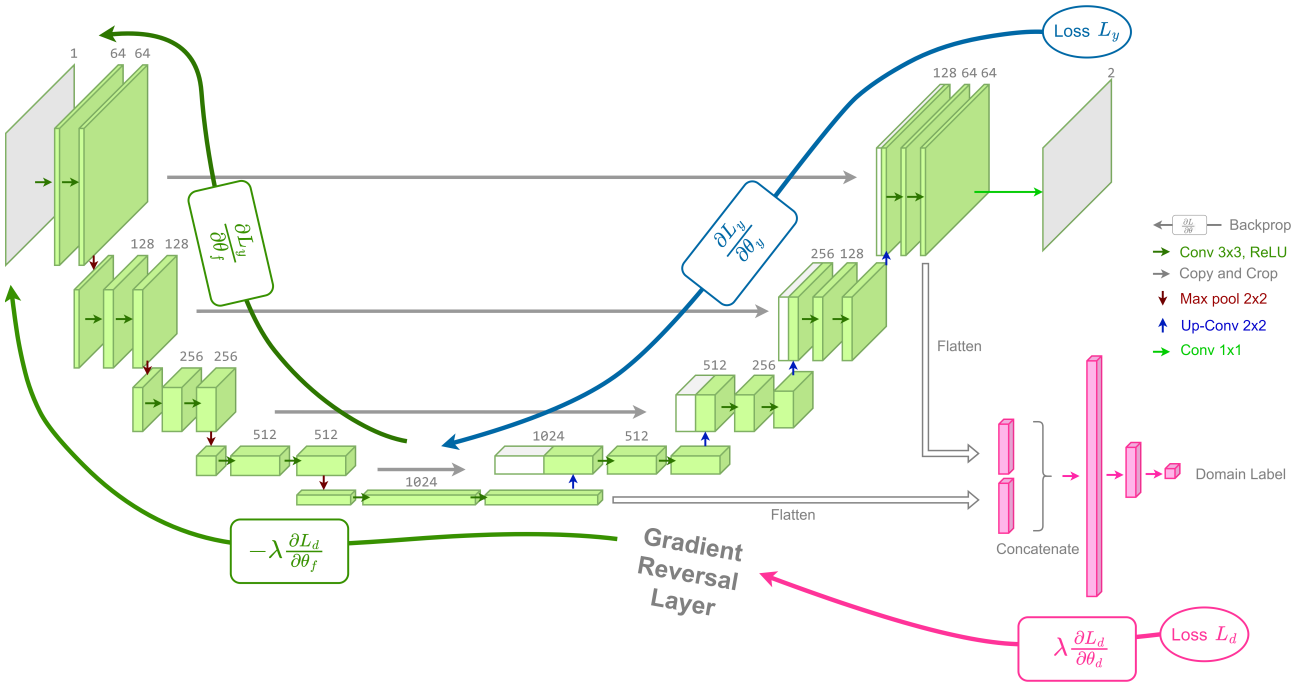


Figure 2: **The overview of our introduced approach.** Our approach consists of two parallel learning loops. The first one is a standard 3D U-net [26] learning loop, where the forward and backward propagation processes are performed on our source images and their corresponding labels. The second loop involves a discriminator. During backpropagation, the discriminator parameters are updated correctly, but the gradients are negated when they reach the encoder, resulting in extracting indistinguishable and more generic features in the feature extractor.

multiplies the gradients by a negative constant, effectively negating them. This leads to the encoder parameters being updated such that it increases the discriminator loss, forcing the feature extractor to learn domain-invariant features.

During training, the model is exposed to labeled adult data with the original and inverted WM-GM contrast and unlabeled newborn data. The contrast inversion is explained in Section 3.2.2. In section 3.2.3, the domain adversarial model is explained in detail.

3.2.2 Contrast Inversion

One major challenge in segmenting newborn brain MRI is the distinct characteristic of immature GM and WM tissue in newborn brains, which exhibit a poorer and reversed contrast compared to adult brains. To mitigate the domain shift caused by this difference in WM-GM contrast, we devise an extra data augmentation that inverts the contrast between WM and GM in adult brain MRIs. This contrast inverted image dataset is then used along with the original images to train our model, enabling successful skull-stripping of both adult and newborn brain MR images.

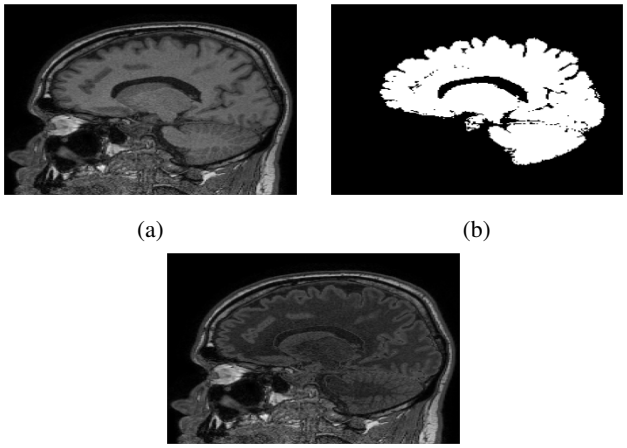
The WM-GM contrast inversion is implemented according to the following equation:

$$\bar{C}(i) = \begin{cases} \max\limits_{\forall j \in M}[C(j)] - C(i), & \forall i \in M \\ C(i), \text{ otherwise} \end{cases} \quad (1)$$

In the above equation, M represents the GM-WM mask, and i denotes all voxels in the mask. By applying the above equation, a new image with reversed contrast denoted by $(\bar{C}(i))$ is created. This new contrast value is obtained by subtracting the original voxel intensity $(C(i))$ from the maximum intensity present in the GM-WM mask, denoted by $(\max[C(j)])$. The contrast inversion is depicted in Fig. 3.

3.2.3 DANN

In our segmentation process, we utilize a modified version of the U-Net model [26] specifically designed for 3D images. The U-Net architecture consists of two main sections: the encoder and the decoder. The encoder follows a 3D convolutional network structure, employing two consecutive $3 \times 3 \times 3$ convolutions (without padding), each followed by a rectified linear unit (ReLU) activation function [13]. Subsequently, a $2 \times 2 \times 2$ max pooling operation with a stride of 2 is applied to downsample the feature channels. At each downsampling step, the number of feature channels is doubled.



(a) (b)

(c)

Figure 3: (a) T1-weighted brain MRI of an adult subject. (B) WM-GM segmentation mask of (a). (C) T1-weighted brain MRI after contrast inversion.

On the other hand, the decoder section of the U-Net performs upsampling of the feature map, followed by a $2\times2\times2$ convolution that reduces the number of feature channels (known as "up-convolution"). The upsampled feature map is then concatenated with the corresponding cropped feature map from the contracting path. Two $3 \times 3 \times 3$ convolutions with ReLU activations are applied to the concatenated feature map. Cropping is necessary to compensate for the loss of border pixels during each convolution. Finally, a $1\times1\times1$ 3D convolution is employed in the last layer to map the 64-component feature vector to the desired number of classes.

In the U-Net architecture, skip connections are established, which directly connect the encoder and decoder sections of the network. These skip connections enable the transfer of information across multiple layers during forward propagation, contributing to the effectiveness of the U-Net in various image segmentation tasks [9] [1].

In our proposed model for DA, we make three major modifications to the traditional U-Net architecture. First, we add a discriminator phase to the model. Second, we concatenate the final layer and the bottleneck of the U-Net and feed them to the discriminator, allowing it to learn both high-level and low-level features. Lastly, we use a gradient reversal layer that reverses the sign of gradients during backpropagation and feeds them to the encoder.

The primary purpose of the discriminator is to differentiate between the source images (adult brain MRI) and the target images (newborn brain MRI). By analyzing the features extracted from the bottleneck layer concatenated to the final layer, the discriminator aims to identify the domain to which an input image belongs.

During the training process, the discriminator is exposed to a combination of source and target images. It learns to classify the images based on their distinctive domain characteristics. The discriminator is equipped with its own set of parameters, which are updated through backpropagation and gradient descent. The objective is to optimize the discriminator's ability to discern between the source and target domains accurately.

After completing the discriminator backpropagation training process, the gradient values are multiplied by a negative coefficient (-λ) through the gradient reversal layer. This manipulation leads to the encoder parameters being updated in a way that increases the discriminator loss, thereby prompting the feature extractor to extract fewer domain-variant features. The intention behind this process is to align the feature distributions of the two domains, making them as similar as possible and indistinguishable for the discriminator. Ultimately, this results in the creation of domain-invariant features that can be effectively used for newborn skull-stripping.

The approach outlined above, which incorporates the gradient reversal layer with a negative coefficient (-λ), proves to be effective for newborn skull-stripping. This technique allows the optimizer to update the model parameters in a manner that promotes the extraction of more generic features. The subsequent equations demonstrate how the introduction of the -λ coefficient modifies the parameter updates, leading to the extraction of features that are less specific to a particular domain and more adaptable to different image characteristics.

$$\theta_y \leftarrow \theta_y - \mu \frac{\partial \mathcal{L}_y^i}{\partial \theta_y} \tag{2}$$

$$\theta_d \leftarrow \theta_d - \mu\lambda \frac{\partial \mathcal{L}_d^i}{\partial \theta_d} \tag{3}$$

$$\theta_f \leftarrow \theta_f - \mu \left(\frac{\partial \mathcal{L}_y^i}{\partial \theta_f} - \lambda \frac{\partial \mathcal{L}_d^i}{\partial \theta_f} \right) \tag{4}$$

The provided equations describe the update process for the parameters (θ) in different components of the model. Equation 2 corresponds to the decoder (θ_y) and demonstrates how its parameters are updated using derivatives from the Dice loss ($\mathcal{L}_y^i$). These derivatives are multiplied by the learning rate (μ) and subtracted from the current decoder parameters to obtain updated values.

Equation 3 pertains to the discriminator (θ_d) and showcases its parameter update. Here, the gradients are obtained from the derivative of the discriminator loss ($\mathcal{L}_d^i$) with respect to the discriminator parameters. The learning rate (μ) and the coefficient (λ) are both multiplied with the gradients to determine the direction and magnitude of the parameter update. The inclusion of the λ term allows the update to be influenced by the domain discrimination objective.

In Equation 4, the update process is applied to the encoder which is the feature detector part(θ_f). Similar to the decoder, chain derivatives from the Dice loss ($\mathcal{L}_y^i$) are utilized. However, in this case, the gradients are modified by subtracting the product of the derivative of the discriminator loss ($\mathcal{L}_d^i$) and the coefficient (λ). The resulting negative λ term in the encoder updating process leads to a reverse update of the encoder parameters.

By incorporating the λ term in the updates, the model focuses on learning features that are challenging for the discriminator to distinguish between the source and target domains. This technique helps in creating domain-invariant features, enhancing the model's ability to perform newborn skull-stripping across different domains.

The λ term is defined as follows:

$$\lambda = \frac{2}{1 + \exp(-10\alpha)} - 1 \tag{5}$$

In equation 5, λ is defined as a coefficient within the gradient reversal layer during the feature extractor's update in the model. To optimize this process, we propose the linking of α to the current position within the data loader across epochs. This approach ensures that α maintains a diminished value in the initial epochs, causing λ to approach proximity to zero. Consequently, during these initial epochs, the feature extractor experiences training through the reverse path of the decoder. This specific training strategy aims at extracting features that hold significance for accurate segmentation. As the progression through the data loader advances, α gradually increases, thereby driving λ closer to 1. This adjustment introduces an additional pathway within the reverse process, originating from the discriminator. This augmentation aids in refining features to be more universally applicable while retaining their importance for segmentation. The underlying rationale for this particular choice of λ is to guarantee that the model gains a robust grasp of pivotal segmentation features before undergoing any domain adaptation. In order to reproduce the results, any coefficient that steers the model's focus towards segmentation features during the initial epochs would suffice.

3.2.4 Loss Function

For the segmentation loss function, we utilize Dice Loss [31], which measures the similarity between the predicted and ground truth segmentation maps by calculating the overlap of corresponding pixels. For the discriminator loss function, we use Binary Cross-Entropy (BCE), which measures the dissimilarity between predicted probabilities and true binary labels. The final loss function used for training this network can be mathematically shown as equation 6.

$$\mathcal{L} = \sum_{i=1}^{N} [\text{Dice}(\hat{\mathbf{y}}_i, \mathbf{y}_i) + \text{BCE}(\mathbf{x}_i, \mathbf{d}_i)] + \sum_{j=1}^{M} \text{BCE}(\mathbf{x}_j, \mathbf{d}_j) \tag{6}$$

In the above equation, the variable $\mathcal{L}$ represents the total loss function. The first summation in the equation consists of two terms. The first term is the Dice Loss applied to the predicted masks ($\hat{\mathbf{y}}_i$) of N source images and the corresponding mask labels ($\mathbf{y}_i$) of them. These values are obtained through training using our 3D U-Net model. The summation over i indicates that we are taking the sum of the Dice Loss over all N source images.

The second term in the first summation is a binary cross-entropy loss applied to the N source images and their corresponding domain labels, represented by $\mathbf{x}_i$ and $\mathbf{d}_i$ respectively. The summation over i indicates that we are taking the sum of the binary cross-entropy loss overall N source images.

The second summation in the equation is also a binary cross-entropy loss, but this time it is applied to M target images and their corresponding domain labels, represented by $\boldsymbol{x}_j$ and $\boldsymbol{d}_j$ respectively. The target images are from a different domain than the source images, and the binary cross-entropy loss is used to help the model adapt to the target domain during D. The summation over j indicates that we are taking the sum of the binary cross-entropy loss overall M target images.

We aim to enhance both the discriminator and segmentation models using loss functions. However, by multiplying -λ in the internal derivation calculations, we influence the encoder to focus more on indistinguishable features, resulting in a continual tug-of-war between the encoder and discriminator. The discriminator loss function is intended to improve its performance, while the features become increasingly difficult to distinguish with each epoch.

3.3. Metrics

In this research, we utilize 95th percentile Hausdorff [16] and Dice coefficient [37], to compare the quality of our brain masks with the manual segmentation reference. The metrics are computed on a slice-by-slice basis. The 95th percentile Hausdorff distance is effective in identifying outliers by capturing the maximum distance between corresponding points on the boundaries. Conversely, the Dice coefficient quantifies segmentation accuracy by measuring the overlap and similarity between the two sets. These widely recognized metrics contribute to a comprehensive evaluation of our brain masks, enabling a robust assessment of segmentation performance.

The 95th percentile Hausdorff distance is a variant of the Hausdorff Distance that measures the maximum distance between two sets of points. In the context of image segmentation, it serves as a valuable metric for evaluating the dissimilarity between the boundaries of the predicted mask and the ground truth mask. This directed distance calculates the maximum distance from each point in one mask to its nearest point in the other mask.

The Dice coefficient is a statistic that assesses the spatial overlap between two binary masks. It calculates the ratio of twice the intersection of the masks to the sum of the sizes of the individual masks. The Dice coefficient ranges from 0 to 1, where a value of 1 indicates a perfect overlap between the masks, while a value of 0 indicates no overlap at all.

4. Experiments

4.1. Experimental Settings

To our knowledge, there are no existing unsupervised models specifically designed for newborn brain skull-stripping in the literature to compare our model directly with. To demonstrate the advancements achieved by our unsupervised model, we conducted a comprehensive comparison involving the current state-of-the-art (SOTA) supervised model for skull-stripping. Among these models, we examine SynthStrip [15], which exhibits promising outcomes in skull-stripping tasks and Hippodeep [32] which was initially developed for extracting the hippocampus from brain MRIs, but its latest PyTorch version has been enhanced to encompass skull-stripping as well. Although both these models are supervised, having been trained on an extensive dataset of MRIs across various age groups, including infants, our model's performance closely rivals their outcomes. This comparison underscores the exceptional capability of our unsupervised model to stand toe-to-toe with established supervised approaches. Also, we performed experiments involving the training and testing of three additional models, alongside our model, on newborn data, allowing for comparative analysis of the results. The first model, referred to as U-Net, was trained solely on adult brain MRI images and subsequently evaluated on newborn data. This model utilized the conventional U-Net architecture without any modifications.

For the second model, denoted as U-Net CI, we trained the same architecture as U-Net but incorporated both the original adult data and the contrast-inverted (CI) adult data during training. The third model, U-Net DA, was trained on the original adult data and featured the inclusion of the discriminator component and gradient reversal layer. Basically, in this model, the architecture is the same as our model, but the model is only fed with the original adult data, not the CI data.

The fourth model was our proposed model, which incorporates both CI and DA elements. The U-Net and the U-net CI models were trained using a Dice loss function, while the U-Net DA and our proposed method were trained using the loss function described in Section 3.2.4.

In all six models, namely U-Net, U-Net CI, U-Net DA, SynthStrip, Hippodeep, and our proposed model, we maintained consistency in terms of the optimizer, learning rate, and dynamic λ. We employed the Adam [18] optimizer with an initial learning rate of $1e{-}4$ for all models. Additionally, to prioritize the early stages of training and enhance segmentation performance, we incorporated a dynamic lambda rate adjustment in all models. This adjustment allowed us to allocate more emphasis and resources toward improving segmentation accuracy during the initial training process.

The training process of all models involved 450 epochs, and it was executed on an A-100 GPU. The chosen batch size for training was 4. Each epoch took approximately 15 minutes to complete on the specified GPU. In order to handle the large volume of data, we adopted a patch-based approach using a patch size of $96 \times 96 \times 96$. This patching technique was facilitated by utilizing the MONAI framework [4]. Due to hardware limitations, we were unable to continuously utilize the GPU. To address this constraint, we implemented a custom function that allowed us to store and retrieve crucial information for resuming the training process. This included saving the model states, optimizer states, the last fully trained epoch, and the corresponding loss function value. This approach ensured that we could resume the training process from the point of interruption and continue the model's progress.

For comparing the models, we computed the metrics described in Section 3.3 using a source domain test set of adult MR images and a target domain test set of newborn brain MR images. The train, validation, and test split used in the experiments are summarized in Table 1.

Table 1: Summary of the train, validation, and test split of the source and target domain datasets used in the experiments. The number of samples of each set corresponds to T1-weighted MRI volumes.

Data	Train Set	Validation Set	Test Set
Adult Data	243	104	12
Newborn Data	5	2	5

4.2. Results

The Dice coefficient results and the 95^{th} percentile Hausdorff distance results for the newborn and adult test sets are summarized in Table 2 and Table 3, respectively. Our proposed method obtained the best metrics in the newborn test set among unsupervised models with a 50% improvement in the Dice coefficient and 35% improvement in the Hausdorff distance metrics compared to the second-best unsupervised method, which was the U-Net DA. Furthermore, in comparison to the cutting-edge supervised models, our model's outcomes stand in remarkable proximity, exhibiting a marginal deviation of 0.7% less than Hippodeep and 2.7% less than SynthStrip in the Dice coefficient. It is worth highlighting that the SynthStrip model has already been trained on a dataset of labeled newborn MRIs, acquired at Boston Children's Hospital from individuals aged between 0 and 18 months, and the Hippodeep model has been trained on an extensive combination of over 5,000 MRI samples.

On the adult test set, our method obtained the second-best Dice coefficient and Hausdorff distance metrics, but the differences in metric quality between the best result and the worst result were smaller than 2%. This underscores the robustness of our model, highlighting that its proficiency in skull-stripping newborn brains does not come with compromising its performance on adult brains.

Segmentation masks representative of the overall results on the newborn and adult test sets are depicted in Fig. 4 and Fig. 5, respectively.

4.3. Discussion

Our innovative unsupervised DA approach to skull-stripping T1-weighted MR images of newborns offers a unique combination of speed and accuracy. By integrating DA techniques, we are able to rapidly generate segmentation results without the need for laborious manual labeling. This streamlined process significantly reduces the overall processing time, enabling healthcare professionals to obtain timely segmentations for clinical analysis. Remarkably, despite being an unsupervised approach, our model delivers reasonably accurate segmentation.

The integration of the CI and DA methods has resulted in significant improvements in the segmentation of newborn brain MRIs

Figure 4: Visual representation of models segmentation results on the newborn test set (target dataset).

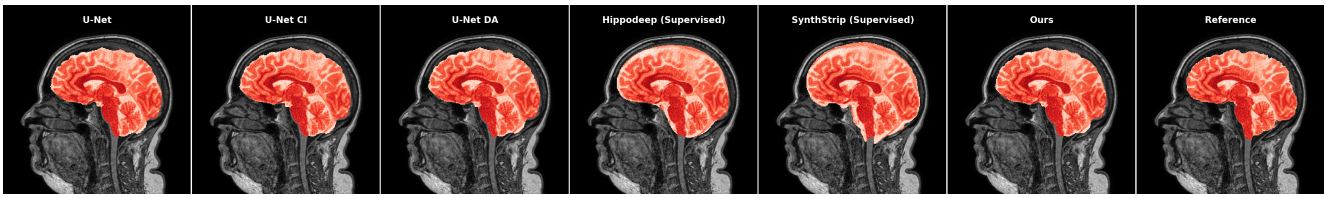


Figure 5: Visual representation of models segmentation results on the adult test set (source dataset).

Table 2: Summary of the results on the newborns' test set.

Approach	Hausdorff Distance	Dice coefficient
U-Net	8.9220 ± 0.5335	0.0790 ± 0.0926
U-Net CI	6.6315 ± 1.0921	0.3863 ± 0.1518
U-Net DA	5.5144 ± 0.7054	0.6086 ± 0.1846
SynthStrip*	**3.1570 ± 0.1389**	**0.9412 ± 0.0063**
Hippodeep*	3.7995 ± 0.2343	0.9222 ± 0.0162
Ours	**3.6621 ± 0.4572**	**0.9156 ± 0.0324**

* Models marked with an asterisk are supervised and already trained on a newborn dataset.

Table 3: Summary of the results on the adults' test set.

Approach	Hausdorff Distance	Dice coefficient
U-Net	**1.6876 ± 0.2778**	0.9513 ± 0.0260
U-Net CI	1.7223 ± 0.2634	0.9500 ± 0.0261
U-Net DA	1.7130 ± 0.2558	0.9514 ± 0.0257
SynthStrip*	2.0116 ± 0.2232	0.9426 ± 0.0165
Hippodeep*	2.0426 ± 0.1973	**0.9576 ± 0.0080**
Ours	**1.6972 ± 0.2687**	**0.9521 ± 0.0262**

* Models marked with an asterisk are supervised and already trained on a newborn dataset.

while maintaining a satisfactory performance on adult brain MRIs. The tables illustrate a remarkable 50% increase in the dice coefficient when compared to using DA alone, and a notable improvement of 137% compared to utilizing CI alone. Notably, there are no existing unsupervised models in the literature for newborn brain skull-stripping to provide a direct comparison. These findings highlight the effectiveness of our method in achieving accurate segmentation and efficient DA, making it a valuable tool in medical imaging applications without the need for manual labels.

Compared with SOTA supervised models, our results are close, nearly 0.7% less than Hippodeep and 2.7% less than SynthStrip. Notably, SynthStrip is trained on newborns labeled MRI from ages 0 to 18 months, and Hippodeep uses over 5000 MRI samples.

However, despite the remarkable accuracy achieved in our results (0.916 ± 0.032), there are some limitations that can serve as potential avenues for future research. To enhance the practical applicability of our model in clinical settings, it is necessary to improve its performance. One potential approach to achieve this is by augmenting the training dataset with a certain amount of labeled newborn data. This additional labeled data can provide valuable guidance and fine-tuning to enhance the model's accuracy.

Furthermore, the current architecture utilized in our research may be relatively large, limiting computational efficiency. To address this concern, we can explore alternative architectures, such as Transformers, which have shown promising results in various domains [14]. By adopting more efficient architectures, we aim to streamline the model and optimize its computational resources while maintaining or even improving its segmentation performance. These future research endeavors will contribute to the overall advancement of our approach and facilitate its practical implementation in real-world scenarios.

5. Conclusions

In conclusion, this research presents a compelling demonstration of our model's ability to adapt to the unique characteristics of MR images and achieve accurate segmentations without reliance on pre-existing labels. These findings hold great significance for medical imaging applications, particularly in situations where obtaining labeled data is challenging.

The high Dice coefficient of 0.9156 ± 0.0324 achieved by us, surpassing other unsupervised approaches examined in the results section, and stands in close proximity to that of the current state-of-the-art supervised models highlighting the substantial potential of our method to impact clinical decision-making. Our findings demonstrate a remarkable 50% increase in the dice coefficient when comparing our model to the use of DA alone and a significant improvement of 137% compared to the use of contrast inversion alone. The improved accuracy in segmentations opens up new possibilities for precise diagnosis and treatment planning, ultimately benefiting patient care.

6. Acknowledgments

Abbas Omidi is a recipient of the Alberta Graduate Excellence Scholarship and extends their gratitude to the organization.

References

[1] Arbish Akram and Nazar Khan. Us-GAN: On the importance of ultimate skip connection for facial expression synthesis. *Multimedia Tools and Applications*, pages 1–17, 2023.

[2] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. *Advances in Neural Information Processing Systems*, 19, 2006.

[3] Benjamin Billot, Douglas N Greve, Oula Puonti, Axel Thielscher, Koen Van Leemput, Bruce Fischl, Adrian V Dalca, Juan Eugenio Iglesias, et al. SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis*, 86:102789, 2023.

[4] M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv preprint arXiv:2211.02701*, 2022.

[5] Hao Chen, Qi Dou, Lequan Yu, Jing Qin, and Pheng-Ann Heng. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage*, 170:446–455, 2018.

[6] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020.

[7] Olivier Colliot, Oscar Camara, and Isabelle Bloch. Integration of fuzzy spatial relations in deformable models—application to brain mri segmentation. *Pattern recognition*, 39(8):1401–1414, 2006.

[8] Alexander de Brebisson and Giovanni Montana. Deep neural networks for anatomical brain segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 20–28, 2015.

[9] Michal Drozdzal, Eugene Vorontsov, Gabriel Chartrand, Samuel Kadoury, and Chris Pal. The importance of skip connections in biomedical image segmentation. In *International Workshop on Deep Learning in Medical Image Analysis, International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis*, pages 179–187. Springer, 2016.

[10] Simon F Eskildsen, Pierrick Coupé, Vladimir Fonov, José V Manjón, Kelvin K Leung, Nicolas Guizard, Shafik N Wassef, Lasse Riis Østergaard, D Louis Collins, Alzheimer's Disease Neuroimaging Initiative, et al. BEaST: brain extraction based on nonlocal segmentation technique. *NeuroImage*, 59(3):2362–2373, 2012.

[11] Yaroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. In *International conference on machine learning*, pages 1180–1189. PMLR, 2015.

[12] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *The journal of machine learning research*, 17(1):2096–2030, 2016.

[13] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 315–323, 2011.

[14] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 574–584, 2022.

[15] Andrew Hoopes, Jocelyn S Mora, Adrian V Dalca, Bruce Fischl, and Malte Hoffmann. Synthstrip: Skull-stripping for any brain image. *NeuroImage*, 260:119474, 2022.

[16] Daniel P Huttenlocher, Gregory A Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, 1993.

[17] Clifford R Jack Jr, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L. Whitwell, Chadwick Ward, et al. The alzheimer's disease neuroimaging initiative (adni): MRI methods. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(4):685–691, 2008.

[18] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[19] Pulkit Kumar, Pravin Nagar, Chetan Arora, and Anubha Gupta. U-segnet: fully convolutional neural network based automated brain tissue segmentation tool. In *2018 25th IEEE International conference on image processing (ICIP)*, pages 3503–3507. IEEE, 2018.

[20] Mingsheng Long, Yue Cao, Jianmin Wang, and Michael Jordan. Learning transferable features with deep adaptation networks. In *International conference on machine learning*, pages 97–105. PMLR, 2015.

[21] Yingli Lu, Tianzi Jiang, and Yufeng Zang. Region growing method for the analysis of functional MRI data. *NeuroImage*, 20(1):455–465, 2003.

[22] Daniel S Marcus, Tracy H Wang, Jamie Parker, John G Csernansky, John C Morris, and Randy L Buckner. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience*, 19(9):1498–1507, 2007.

[23] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[24] Kenichi Oishi, Andreia V Faria, and Susumu Mori. Advanced neonatal neuroMRI. *Magnetic Resonance Imaging Clinics*, 20(1):81–91, 2012.

[25] Sergio Pereira, Americo Oliveira, Victor Alves, and Carlos A Silva. On hierarchical brain tumor segmentation in mri using fully convolutional neural networks: a preliminary study. In *2017 IEEE 5th Portuguese meeting on bioengineering (ENBENG)*, pages 1–4. IEEE, 2017.

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, pages 234–241. Springer, 2015.

[27] Parisa Saat, Nikita Nogovitsyn, Muhammad Yusuf Hassan, Muhammad Athar Ganaie, Roberto Souza, and Hadi Hemmati. A Domain Adaptation Benchmark for T1-weighted Brain Magnetic Resonance Image Segmentation. *Frontiers in Neuroinformatics*, page 96, 2022.

[28] David W Shattuck, Mubeena Mirza, Vitria Adisetiyo, Cornelius Hojatkashani, Georges Salamon, Katherine L Narr, Russell A Poldrack, Robert M Bilder, and Arthur W Toga. Construction of a 3D probabilistic atlas of human cortical structures. *Neuroimage*, 39(3):1064–1080, 2008.

[29] Feng Shi, Pew-Thian Yap, Guorong Wu, Hongjun Jia, John H Gilmore, Weili Lin, and Dinggang Shen. Infant brain atlases from neonates to 1- and 2-year-olds. *PLOS ONE*, 6(4):e18746, 2011.

[30] Roberto Souza, Oeslle Lucena, Julia Garrafa, David Gobbi, Marina Saluzzi, Simone Appenzeller, Letícia Rittner, Richard Frayne, and Roberto Lotufo. An open, multi-vendor, multi-field-strength brain MR dataset and analysis of publicly available skull stripping methods agreement. *NeuroImage*, 170:482–494, 2018.

[31] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017.

[32] Benjamin Thyreau, Kazunori Sato, Hiroshi Fukuda, and Yasuyuki Taki. Segmentation of the hippocampus by transferring algorithmic knowledge for large cohort processing. *Medical image analysis*, 43:214–228, 2018.

[33] Eric Tzeng, Judy Hoffman, Trevor Darrell, and Kate Saenko. Simultaneous deep transfer across domains and tasks. In *Proceedings of the IEEE international conference on computer vision*, pages 4068–4076, 2015.

[34] David C Van Essen, Kamil Ugurbil, Edward Auerbach, Deanna Barch, Timothy EJ Behrens, Richard Bucholz, Acer Chang, Liyong Chen, Maurizio Corbetta, Sandra W Curtiss, et al. The human connectome project: a data acquisition perspective. *NeuroImage*, 62(4):2222–2231, 2012.

[35] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. *Neurocomputing*, 312:135–153, 2018.

[36] Neil I Weisenfeld and Simon K Warfield. Automatic segmentation of newborn brain MRI. *NeuroImage*, 47(2):564–572, 2009.

[37] Kelly H Zou, Simon K Warfield, Aditya Bharatha, Clare MC Tempany, Michael R Kaus, Steven J Haker, William M Wells III, Ferenc A Jolesz, and Ron Kikinis. Statistical validation of image segmentation quality based on a spatial overlap index1: scientific reports. *Academic Radiology*, 11(2):178–189, 2004.