



Automatic Skull Stripping of Rat and Mouse Brain MRI Data Using U-Net

Li-Ming Hsu^{1,2,3,4}, Shuai Wang^{2,4}, Paridhi Ranadive¹, Woomi Ban^{1,2},
Tzu-Hao Harry Chao^{1,2,3}, Sheng Song^{1,2,3}, Domenic Hayden Cerri^{1,2,3},
Lindsay R. Walton^{1,2,3}, Margaret A. Broadwater^{1,2,3}, Sung-Ho Lee^{1,2,3}, Dinggang Shen^{2,4,5*}
and Yen-Yu Ian Shih^{1,2,3*}

¹ Center for Animal Magnetic Resonance Imaging, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, ² Biomedical Research Imaging Center, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, ³ Department of Neurology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, ⁴ Department of Radiology, The University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, ⁵ Department of Brain and Cognitive Engineering, Korea University, Seoul, South Korea

OPEN ACCESS

Edited by:

Anand Joshi,
University of Southern California, Los Angeles, United States

Reviewed by:

John Soraghan,
University of Strathclyde,
United Kingdom
Vladimir S. Fonov,
McGill University, Canada
Xiaoying Tang,
Southern University of Science
and Technology, China

*Correspondence:

Yen-Yu Ian Shih
shihy@unc.edu
Dinggang Shen
dinggang_shen@med.unc.edu

Specialty section:

This article was submitted to
Brain Imaging Methods,
a section of the journal
Frontiers in Neuroscience

Received: 01 June 2020

Accepted: 13 August 2020

Published: 07 October 2020

Citation:

Hsu L-M, Wang S, Ranadive P,
Ban W, Chao T-HH, Song S,
Cerri DH, Walton LR, Broadwater MA,
Lee S-H, Shen D and Shih Y-YI (2020)
Automatic Skull Stripping of Rat
and Mouse Brain MRI Data Using
U-Net. *Front. Neurosci.* 14:568614.
doi: 10.3389/fnins.2020.568614

Accurate removal of magnetic resonance imaging (MRI) signal outside the brain, a.k.a., skull stripping, is a key step in the brain image pre-processing pipelines. In rodents, this is mostly achieved by manually editing a brain mask, which is time-consuming and operator dependent. Automating this step is particularly challenging in rodents as compared to humans, because of differences in brain/scalp tissue geometry, image resolution with respect to brain-scalp distance, and tissue contrast around the skull. In this study, we proposed a deep-learning-based framework, U-Net, to automatically identify the rodent brain boundaries in MR images. The U-Net method is robust against inter-subject variability and eliminates operator dependence. To benchmark the efficiency of this method, we trained and validated our model using both in-house collected and publicly available datasets. In comparison to current state-of-the-art methods, our approach achieved superior averaged Dice similarity coefficient to ground truth T2-weighted rapid acquisition with relaxation enhancement and T2*-weighted echo planar imaging data in both rats and mice (all $p < 0.05$), demonstrating robust performance of our approach across various MRI protocols.

Keywords: rat brain, mouse brain, MRI, U-net, segmentation, skull stripping, brain mask

INTRODUCTION

Magnetic resonance imaging (MRI) is a widely employed technique to study brain anatomy and function in preclinical rodent models (Mandino et al., 2020). To achieve individual subject data standardization and facilitate group level comparison, pre-processing must remove non-brain tissue, a.k.a. skull strip; without it, the automatic registration process would likely fail due to unwanted signal outside the brain. In many cases, skull stripping is achieved by manually drawing brain masks for each individual slice, making it a time-consuming and operator-dependent process. Ideally, an automatic skull stripping tool would streamline the pre-processing pipeline, avoid personnel bias, and significantly improve research efficiency, especially when handling large datasets (Babalola et al., 2009; Lu et al., 2010; Gaser et al., 2012; Feo and Giove, 2019). In human MRI research, several automatic brain extraction tools have been developed and widely utilized (Cox, 1996; Shattuck and Leahy, 2002; Leung et al., 2011; Doshi et al., 2013). However,

these tools are not applicable to rodent applications because of differences in brain/scalp tissue geometry, image resolution with respect to brain-scalp distance, tissue contrast around the skull, and sometimes signal artifacts from surgical manipulations. Additionally, rodent brain MRI data is typically acquired at higher magnetic fields (mostly $>7T$) with higher radiofrequency (RF) coil inhomogeneity. The stronger susceptibility artifacts and field biases represent further challenges to the rodent skull stripping process.

To date, several attempts have been made to address rodent skull-segmentation (Pfefferbaum et al., 2004; Sharief et al., 2008; Bendazzoli et al., 2019; Feo and Giove, 2019; Lohmeier et al., 2019; Liu et al., 2020). To date, the most prominent tools for rodent MRI skull stripping are Pulse-Coupled Neural Network (PCNN)-based brain extraction proposed by Chou et al. (2011), Rapid Automatic Tissue Segmentation (RATS) pioneered by Oguz et al. (2014), and, and SHape descriptor selected External Regions after Morphologically filtering (SHERM) by Liu et al. (2020). Pulse-Coupled Neural Network is a biomimetic neural network initially developed for cat visual cortex segmentation (Kuntimad and Ranganath, 1999) that utilizes an iterative process to assign labels to adjacent pixels with similar intensity profiles. The RATS technique is built on mathematical morphology and LOGISMOS-based graph segmentation methods (Yin et al., 2010). While the RATS method has superior performance on T1-weighted images (T1w; Oguz et al., 2014), it is worth noting that T2-weighted images (T2w) and T2*-weighted images (T2*w) are also common choices in high-field brain function studies. The recently proposed SHape descriptor SHERM (Liu et al., 2020) method identifies a set of brain mask candidates, extracted from MRI images with multiple kernel sizes that matches the shape of the brain template. One common limitation of these brain segmentation methods is that the performance varies by brain size, shape, texture, and contrast, and therefore the technique needs to be optimized for each MRI protocol. Taken together, the development of a rodent skull stripping tool capable of performing on a variety of datatypes with accuracy and consistency is highly desirable.

Instead of using rules designed by users, learning-based methods acquire mapping functions from inbuilt feature engineering and classifiers, which would likely be more robust to various imaging modalities. Specifically, deep-learning-based methods combine feature engineering and classifiers into a uniform framework, and have achieved outstanding performance on many medical imaging identification tasks (Kleesiek et al., 2016; Havaei et al., 2017; Roy et al., 2018). Here we propose a novel model that adopts a fully convolutional deep-learning network, U-Net (Ronneberger et al., 2015; Yogananda et al., 2019), to perform dense feature extraction. The whole network is implemented using Keras (Chollet, 2015) with TensorFlow (Abadi et al., 2016) as the backend. We trained and tested the U-Net model for skull stripping performance using rat and mouse datasets that contained different imaging contrasts [i.e., T2w rapid acquisition with relaxation enhancement (T2w RARE) and T2*w using echo planar imaging (T2w EPI)]. The performance of our proposed model was then compared with

existing rodent skull stripping tools, including RATS, PCNN, and SHERM across different available datasets.

MATERIALS AND METHODS

Dataset Descriptions

This study includes two separate datasets: an in-house collected dataset (CAMRI dataset) and an open source dataset (Online dataset) downloaded from <http://openneuro.org>. The CAMRI dataset consisted of 132 adult male rats of different strains [94 Sprague Dawley (SD), 22 Long-Evans (LE), and 16 Wistar: this dataset is available at <https://doi.org/10.18112/openneuro.ds002870.v1.0.0>] and 16 wild-type adult C57Bl/6J strain mice (the dataset is available at <https://doi.org/10.18112/openneuro.ds002868.v1.0.0>). For each animal, a T2w RARE and an T2*w EPI were acquired. Among the 132 rats, 69 rats' T2w RARE and T2*w EPI resolutions were 0.1 mm \times 0.1 mm \times 1 mm and 0.32 mm \times 0.32 mm \times 1 mm, respectively, and the other 63 rats' T2w RARE and T2*w EPI resolutions were 0.2 mm isotropic and 0.4 mm isotropic, respectively. For the mice, the T2WI and T2*w EPI resolutions were 0.16 mm isotropic and 0.32 mm isotropic, respectively. All CAMRI data were acquired on a Bruker 9.4T system. The Online dataset consisted of 24 rats and 36 mice. Specifically, T2w RARE of 24 female adult Wistar strain rats (Sirmipilatzte et al., 2019),¹ T2w RARE of 16 male and female B6.Cg-Tg(Fev-cre)1Esd/J mice (ePet-cre; RRID:IMSR_JAX:012712) (Grandjean et al., 2019),² and T2*w EPI images of 20 C57Bl/6J male and female mice (Grandjean et al., 2020).³ To train our U-Net model, we first established training dataset by randomly selecting 80% of the T2w RARE and T2*w EPI images in the CAMRI rat data (78 SD, 15 LE, and 12 Wistar) as well as all CAMRI mouse data, leaving the remaining 20% of the rats as final performance testing dataset. In the training process, we further randomly selected 80% of the rat data from the training dataset (62 SD, 12 LE, and 10 Wistar) and included all mouse data for inner training. The remaining 20% of the rat data from the training dataset was used to validate the U-Net model. We repeated the training-validation process five times to avoid randomness in the data splitting. The U-Net model with the highest averaged validation accuracy was then used as the final model for testing.

To further illustrate the robustness and wide applicability of our proposed model in separate rat and mouse datasets, we tested our trained U-Net model on the Online dataset that was acquired from different scanners and with different imaging parameters.

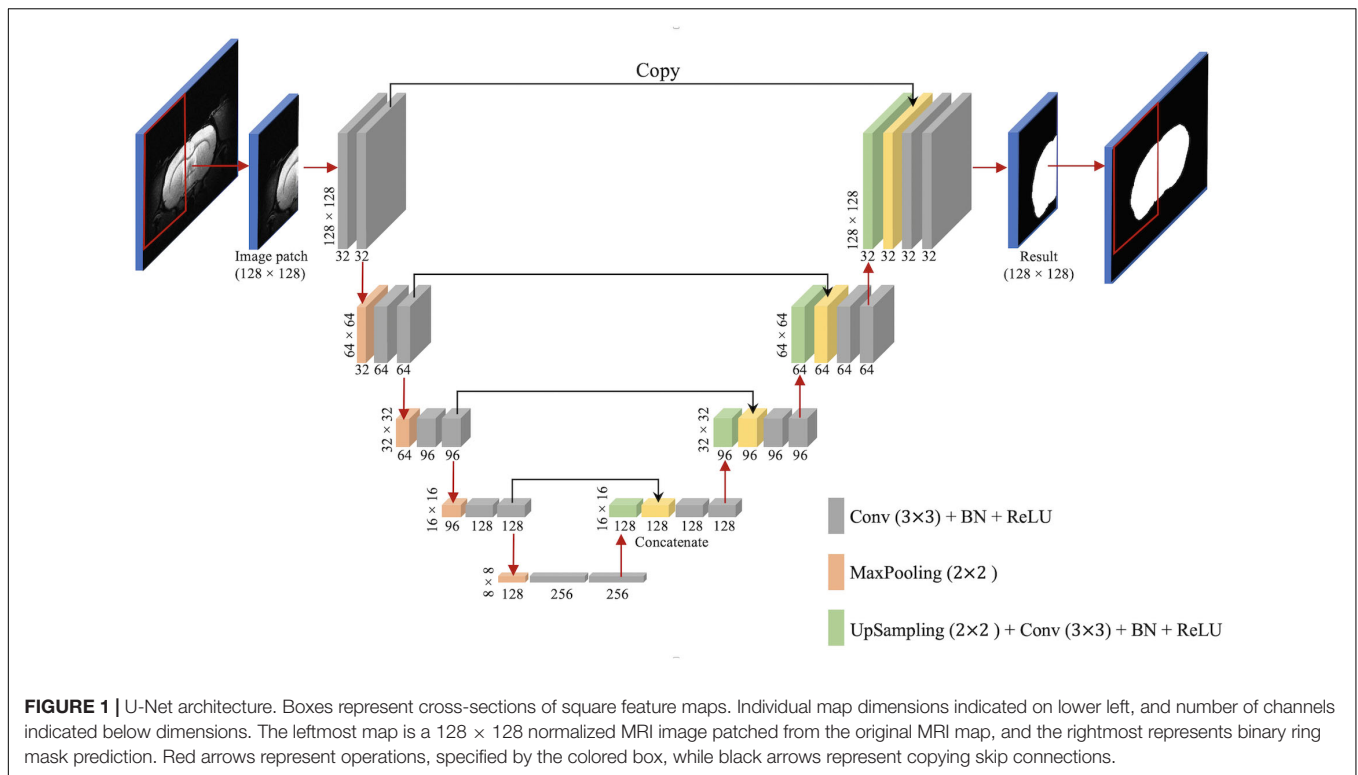
U-Net

We used U-Net (Ronneberger et al., 2015), a method with excellent performance in many medical image segmentation tasks (Ronneberger et al., 2015; Zhou et al., 2018; Alom et al., 2019; Yogananda et al., 2019; Wang et al., 2020b), to perform skull stripping on rodent brain MR images (Figure 1). In

¹<https://doi.org/10.18112/openneuro.ds001981.v1.0.3>

²<https://doi.org/10.18112/openneuro.ds001541.v1.1.2>

³<https://doi.org/10.18112/openneuro.ds001720.v1.0.2>



the contracting path, there are 32 feature maps in the first convolutional block, 64 in the second, then 96, 128, and 256 in the third, fourth, and fifth, respectively. Compared to the configuration described by Ronneberger et al. (2015), we replaced the cross-entropy loss function with the Dice coefficient loss (Wang et al., 2020a) to free the optimization process from a class-imbalance problem (Milletari et al., 2016).

In this study, since we include various rats and mice dataset (CAMRI and online dataset) with different image resolutions, we performed two different normalizations to improve the capabilities of the model: spatial normalization and intensity normalization. For spatial normalization, we resampled all images into the same spatial resolution at $0.1 \text{ mm} \times 0.1 \text{ mm}$ slice-by-slice using nearest-neighbor interpolation. The nearest-neighbor interpolation was chosen to keep consistency in the processing pipeline because both brain-mask (binary) and brain image (grayscale) need to be resampled. Resampling was not performed across slices because we performed 2D U-Net slice-by-slice. For intensity normalization, we performed the min-max normalization for each image to range intensities from 0 to 1 and stored them as single precision (float-32). In U-Net training, the voxels belonging to the rat brain are labeled as 1 and other voxels (background) are labeled as 0. Our network was implemented using Keras (Chollet, 2015) with TensorFlow (Abadi et al., 2016) as the backend. The initial learning rate and batch size were $1e^{-3}$ and 16, respectively. We used Adam (Kingma and Ba, 2015) as the optimizer and clipped all parameter gradients to a maximum norm of 1. In training, we randomly cropped the 128×128 sized patches from all axial slices as the input. In the inference, the overlapped patches extracted from each axial slice

were input into the trained model with a $16 \times 16 \times 1$ stride. The overlapped predictions were averaged and then resampled back to the original resolution using nearest-neighbor interpolation for the final output.

Evaluation Methods

To demonstrate the reliability of our proposed method, we compared our U-Net method with the most prominently used methods for rat brain segmentation: RATS (Oguz et al., 2014), PCNN (Chou et al., 2011), and SHERM (Liu et al., 2020). All images were bias-corrected for field inhomogeneities using Advanced Normalization Tools (ANTs).⁴ Since we included multiple datasets in this study, the parameters were chosen according to best parameters suggested in the publication to maintain consistency. For the RATS algorithm, the intensity threshold (T) was set to the average intensity in the entire image and the brain size values V_t was set to 1650 mm^3 for the rat images and 380 mm^3 for mouse images (Oguz et al., 2014). For the PCNN algorithm, the brain size range was set to $1000\text{--}3000 \text{ mm}^3$ for rat images and $350\text{--}550 \text{ mm}^3$ for mouse images (Chou et al., 2011). For SHERM, the brain size range was set to $500\text{--}1900 \text{ mm}^3$ for rat images and $300\text{--}550 \text{ mm}^3$ for mouse images (Liu et al., 2020). The default convexity threshold in SHERM, defined as the ratio between the volume of a region and that of its convex hull, was set to 0.85 to discard brain mask candidates. We adjusted the convexity threshold to 0.7 because brain mask candidate did not survive in half of the rodent images

⁴<http://stnava.github.io/ANTs/>

from CAMRI and online datasets, likely due to differences in raw data dimensions.

To quantitatively evaluate the segmentation performance of U-Net, RATS, PCNN, and SHERM, we estimated the similarity of the brain segmentation results generated by each method compared to manual drawing of brain masks by an anatomical expert according to the Paxinos and Watson rat atlas (Paxinos and Watson, 2014) and Kõnsman mouse atlas (Kõnsman, 2003). The manual segmentation was performed at the original MRI resolution before data resampling to $0.1 \text{ mm} \times 0.1 \text{ mm}$ for U-Net training. To evaluate the reliability of the manual delineations (ground truth), we included two additional experts with profound knowledge of rodent brain anatomy and estimated the inter-rater accuracy compared to ground truth using 20 randomly selected rats (both T2w RARE and T2*w EPI images). High reliability (accuracy > 0.95 , **Supplementary Figure 2**) of the ground truth was found. Evaluations included: (1) volumetric overlap assessments via Dice, the similarity of two samples; (2) Jaccard, the similarity of two samples where Dice doesn't satisfy the triangle inequality; (3) positive predictive value (PPV), the rate of true positives in prediction results; and (4) sensitivity (SEN), the rate of true positives in manual delineation; as well as (5) a surface distance assessment by Hausdorff distance, the distance of two samples. The following definitions were used for each: $Dice = 2(|A \cap B|)/(|A| + |B|)$, $Jaccard = (|A \cap B|)/(|A \cup B|)$, $PPV = (|A \cap B|)/B$, $SEN = (|A \cap B|)/A$, and $Hausdorff = \max \{h(A, B), h(B, A)\}$

and $h(A, B) = \max \{ \min d(a, b) \}$ where A denotes the voxel $a \in A$ $b \in B$

set of the manually delineated volume, B denotes the voxel set of the predicted volume, and $d(a, b)$ as the Euclidian distance between a and b . The Hausdorff distance was only estimated in-plane to avoid confounds from non-uniformly sampled data. The maximal Hausdorff distance (i.e., worst matching) across slices for each subject was then used for comparison. Superior performance was indicated by higher Dice, Jaccard, PPV, and SEN, and lower Hausdorff values. We also reported the computation time on a Linux-based [Red Hat Enterprise Linux Server release 7.4 (Maipo)] computing system (Intel E5-2680 v3 processor, 2.50 GHz, 256-GB RAM) for each method. The computation times reported do not include any preprocessing steps (i.e., signal normalization, image resampling, and bias correction). Paired t-tests were used for statistical comparisons between different algorithms, and two-sample t-tests were used to compare T2w RARE and T2*w EPI images in each algorithm. The threshold for significance was set to the alpha level ($p < 0.05$).

RESULTS

Figure 2 illustrates the performance of our trained U-Net algorithm compared to RATS (Oguz et al., 2014) and PCNN (Chou et al., 2011) for rat brain segmentation in the CAMRI dataset. In all measures, U-Net performed significantly better than the other two methods, except PPV was slightly inferior to

RATS on the T2*w EPI dataset. Notably, U-Net produced near-perfect results with all measures of volumetric overlap > 0.90 . In contrast, the high PPV (0.98 on T2w RARE and 0.99 on T2*w EPI) but low SEN (0.85 on T2w RARE and 0.75 on T2*w EPI) from RATS indicates segmentation was underestimated, while the low PPV (0.85 on T2w RARE and 0.72 on T2*w EPI) and high SEN (0.90 on T2w RARE and 0.93 on T2*w EPI) in PCNN indicates segmentation was overestimated. The significantly lower Hausdorff distance in U-Net (4.27 on anisotropic T2w RARE and 4.60 on anisotropic T2*w EPI) further indicates its best match segmentation. However, the U-Net algorithm had longer computation time than others using the same computational environment (67.66 s on T2w RARE and 64.70 s on T2*w EPI). In summary, the high accuracy (Dice > 0.95) of U-Net in training, validating (**Supplementary Figure 1**), and final performance testing demonstrates the reliability and consistency of our method.

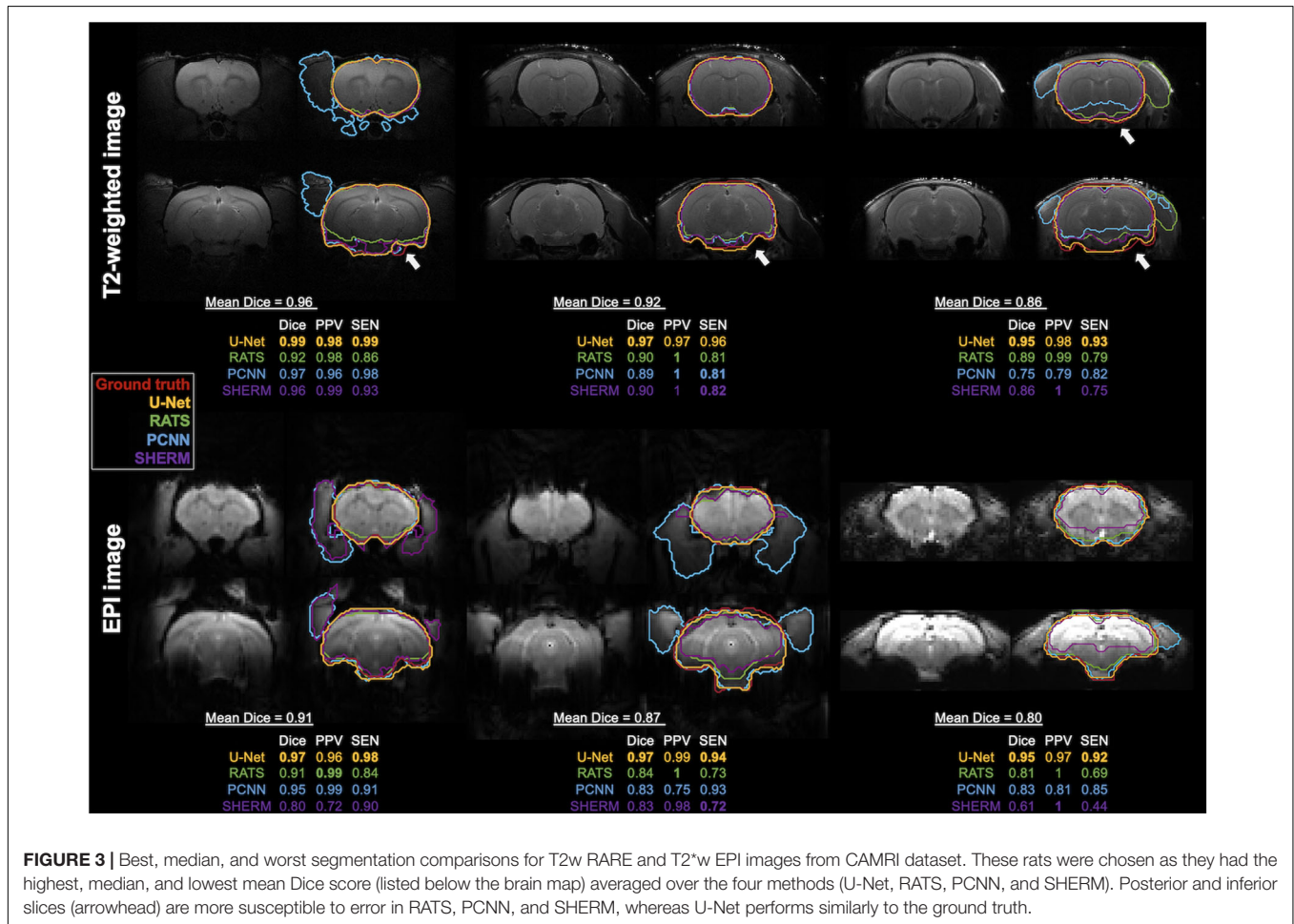
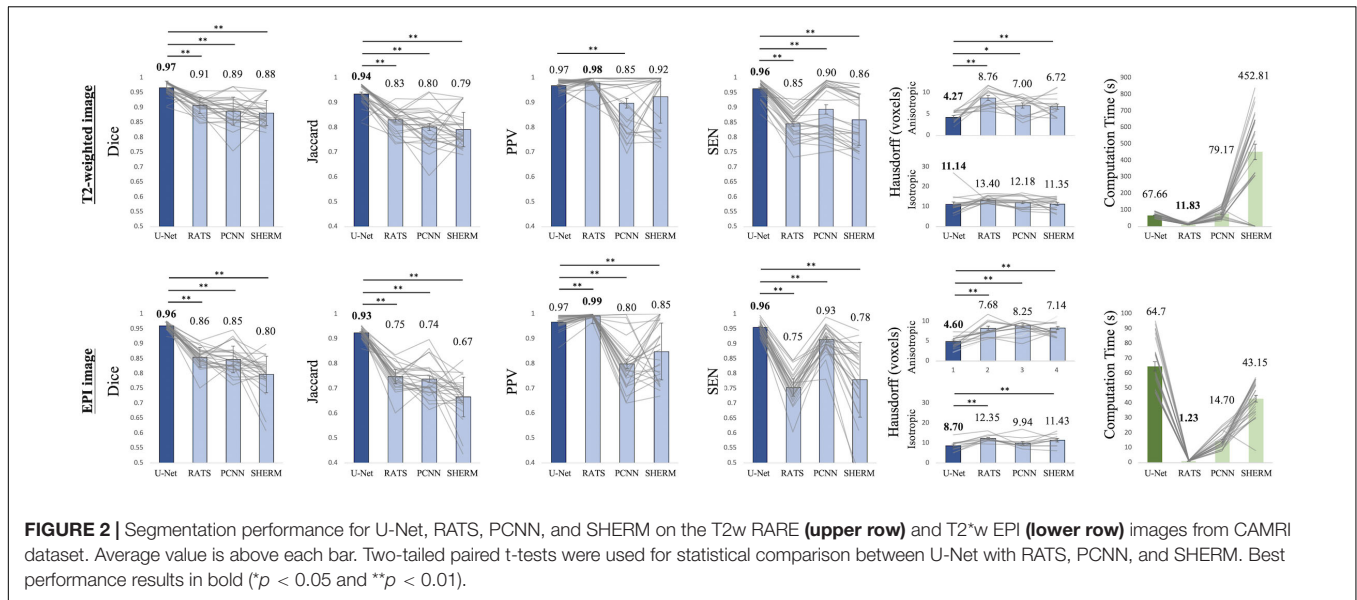
There were no significant differences in segmentation performance between T2w RARE and T2*w EPI with U-Net, but a significant decrease in performance was found with the other three algorithms (All $p < 0.05$, **Figure 2**). Specifically, the Dice, Jaccard, PPV, and SEN from RATS, the Dice, Jaccard, and PPV from PCNN, and the Dice, Jaccard, PPV, and SEN from SHERM were lower for T2*w EPI than T2w RARE. The compromised performance in the T2*w EPI image compared with T2w RARE indicates the challenges these three methods have with low resolution images.

Figure 3 illustrates the best, median, and worst cases on T2w RARE and T2*w EPI from the CAMRI dataset using all four algorithms. These chosen rats had the highest, median, and lowest Dice score averages over the four methods. Note that in the worst case the RATS, PCNN, and SHERM algorithms failed to identify the brainstem, olfactory bulb, and inferior brain regions where the MRI signal was weaker. **Supplementary Figure 3** illustrates more results for T2*w EPI images. Importantly, U-Net could still achieve a satisfactory segmentation in the worst cases with Dice > 0.95 for both T2w RARE and T2*w EPI. Compromised MRI signal intensity causes problems for RATS, PCNN, and SHERM algorithms, while U-Net still produces near-perfect results.

We included the Online dataset to illustrate the performance of our proposed algorithm on independent rat and mouse datasets. **Table 1** indicates segmentation performance for rat T2w RARE. U-Net performed significantly better than RATS, PCNN, and SHERM on nearly all measures except PPV. Both T2w RARE (**Table 2**) and T2*w EPI (**Table 3**) skull stripping in the mouse dataset were significantly improved in U-Net versus the other two methods except for PPV and Hausdorff distance. Overall, these results indicate that the proposed U-Net method is a highly competitive alternative to other existing skull stripping tools.

DISCUSSION

Our results indicate that our proposed skull stripping framework based on U-Net represents a robust method for the accurate and automatic extraction of rodent brain tissue



from MR images. While existing rodent skull stripping methods are robust when used with high-resolution anatomical images, most of them face challenges with low

resolution, low contrast T2*w EPI datasets. Overall, the U-Net based method showed consistent performance in both T2w RARE and T2*w EPI, likely attributed to the

TABLE 1 | Quantitative comparison of U-Net, RATS, PCNN, and SHERM for segmentations on rat T2w RARE from Online dataset.

Methods	Dice	Jaccard	PPV	SEN	Hausdorff (voxels)
U-Net	0.94 (0.00)	0.88 (0.01)	0.94 (0.00)	0.98 (0.01)	6.81 (0.44)
RATS	0.89 (0.02)	0.82 (0.03)	0.95 (0.01)	0.86 (0.02)	8.38 (0.40)
PCNN	0.85 (0.02)	0.75 (0.03)	0.84 (0.03)	0.88 (0.02)	9.16 (0.91)
SHERM	0.85 (0.02)	0.75 (0.02)	0.95 (0.01)	0.78 (0.03)	9.81 (0.88)
<i>p</i> -value (U-Net vs. RATS)	<0.05	<0.05	N.S.	<0.05	<0.005
<i>p</i> -value (U-Net vs. PCNN)	<0.001	<0.001	<0.005	<0.005	<0.05
<i>p</i> -value (U-Net vs. SHERM)	<0.001	<0.001	N.S.	<0.001	<0.005

The *p*-value indicates the result of two-tailed paired *t*-test comparison (best performance results in bold).

TABLE 2 | Quantitative comparison of U-Net, RATS, PCNN, and SHERM for segmentations on mouse T2w RARE from Online dataset.

Methods	Dice	Jaccard	PPV	SEN	Hausdorff (voxels)
U-Net	0.85 (0.01)	0.74 (0.01)	0.74 (0.01)	0.98 (0.00)	5.23 (0.37)
RATS	0.82 (0.01)	0.70 (0.01)	0.76 (0.01)	0.91 (0.01)	5.07 (0.31)
PCNN	0.79 (0.00)	0.65 (0.01)	0.76 (0.01)	0.83 (0.01)	7.07 (0.47)
SHERM	0.80 (0.01)	0.67 (0.01)	0.72 (0.01)	0.90 (0.01)	7.03 (0.41)
<i>p</i> -value (U-Net vs. RATS)	<0.05	<0.05	N.S.	<0.001	N.S.
<i>p</i> -value (U-Net vs. PCNN)	<0.001	<0.001	N.S.	<0.001	<0.005
<i>p</i> -value (U-Net vs. SHERM)	<0.001	<0.001	N.S.	<0.001	<0.005

The *p*-value indicates the result of two-tailed paired *t*-test comparison (best performance results in bold).

TABLE 3 | Quantitative comparison of U-Net, RATS, PCNN, and SHERM for segmentations on mouse T2*w EPI images from Online dataset.

Methods	Dice	Jaccard	PPV	SEN	Hausdorff (voxels)
U-Net	0.92 (0.01)	0.85 (0.01)	0.91 (0.01)	0.93 (0.01)	3.57 (0.14)
RATS	0.85 (0.01)	0.75 (0.01)	0.97 (0.00)	0.76 (0.01)	3.85 (0.11)
PCNN	0.87 (0.01)	0.77 (0.01)	0.86 (0.01)	0.88 (0.01)	3.79 (0.16)
SHERM	0.87 (0.01)	0.77 (0.01)	0.92 (0.01)	0.82 (0.01)	3.39 (0.10)
<i>p</i> -value (U-Net vs. RATS)	<0.001	<0.001	<0.001	<0.001	N.S.
<i>p</i> -value (U-Net vs. PCNN)	<0.001	<0.001	<0.005	<0.001	N.S.
<i>p</i> -value (U-Net vs. SHERM)	<0.001	<0.001	N.S.	<0.001	N.S.

The *p*-value indicates the result of two-tailed paired *t*-test comparison (best performance results in bold).

use of both T2w RARE and T2*w EPI images to train our U-Net architecture.

Compared to the pioneering techniques RATS (Oguz et al., 2014), PCNN (Chou et al., 2011), and SHERM (Liu et al., 2020), our proposed U-Net architecture is more robust, likely due to its capability to explore and learn the hierarchical features from the training dataset without requiring additional parameter adjustments. U-Net combines the location information from the downsampling path with the contextual information in the upsampling path to obtain a combination of localization and contextualization necessary to predict a reliable segmentation (Ronneberger et al., 2015). One clear advantage of the U-Net algorithm is that it is parameter free in the segmentation process, as all parameters are automatically learned from the data itself. The only parameters to learn on convolution layers in U-Net are the kernel. The size of the kernel is independent from the input image size, so images of different sizes can be used as input. In contrast, both RATS and PCNN need to select the appropriate brain size for rat or mouse brain for accurate

justification. In RATS, the intensity threshold also needs to be adjusted to remove low signal intensity as potential non-brain signal. In practice, users need to adjust these parameters once per study based on the acquisition protocol, which affects the intensity profile, and the age/species/strain of the animals, which affects expected brain sizes. Note that, RATS, PCNN, and SHERM still reach an accurate (Dice > 0.8) and fast segmentation performance whereas the U-Net architecture requires longer processing time and needs a higher level of computational power for architecture training. Typically, deep learning-based methods are time-consuming in central processing units (CPU) but are significantly more time-efficient in graphics processing units (GPU). Indeed, the computation time of our proposed U-Net application can benefit significantly by using GPU (**Supplementary Figure 4**). Besides, conventional rodent brain extraction algorithms were based on prior knowledge of rodent brain anatomy, or adapting a general-purpose segmentation method, so an image covering the complete rat brain is necessary for basic functioning. In contrast, since the U-Net architecture

learns the features for each slice, it could still work with images covering a limited brain section.

The robustness of U-Net is clearly illustrated in the segmentation performance of selected-cases across different protocols. Due to relatively poor signal intensity in the brainstem, olfactory bulb, and inferior part of brain, RATS, PCNN, and SHERM displayed lower segmentation accuracy in these areas in T2w RARE and T2*w EPI. Although all methods provided outstanding segmentation performance (Dice > 0.9), the best T2w RARE and T2*w EPI segmentation comparisons still showed mismatches in the inferior part of brain in RATS, PCNN, and SHERM. Furthermore, outcome assessments using different MRI protocols (T2w RARE and T2*w EPI images) indicate that U-Net has high accuracy and consistency across various resolutions. Notably, while most brain segmentation was performed in the anatomical image (T2w RARE), our proposed U-Net architecture also shows accuracy in the T2*w EPI images. When comparing the skull stripping results between T2w RARE and T2*w EPI images in the CAMRI dataset, PCNN, RATS, and SHERM showed significantly lower segmentation accuracy in T2*w EPI images while no significant difference was displayed in the U-Net algorithm. Specifically, in the worst case of T2w RARE image (**Figure 3**), the RATS displayed PPV = 0.99 and SEN = 0.79, which indicated the identified brain tissue has a high rate of true positive but low rate of false negative predictions, and the opposite performance was found in PCNN (PPV = 0.79 and SEN = 0.82). A similar trend was also found in the worst case of T2*w EPI image (**Figure 3**). The T2*w EPI outcome in RATS is underestimated and in PCNN is overestimated, which makes U-Net the superior choice for skull stripping these lower resolution images (PPV = 0.99 and SEN = 0.94). We observed the similar skull stripping performance for T2 W (Dice = 0.97) and EPI (Dice = 0.96), indicating that the model is adequately trained and not susceptible to ghosting artifacts in EPI. Rodent EPI data is also less prone to motion because the subjects are either under anesthesia and secured with ear and tooth bars (Atay et al., 2008; Albaugh et al., 2016; Van Den Berge et al., 2017; Broadwater et al., 2018; Grandjean et al., 2019, 2020; Sirmipilatzte et al., 2019; Mandino et al., 2020) or awake and tightly restrained (Madularu et al., 2017; Ma et al., 2018). Indeed, none of the dataset available on online repository suffers from severe EPI ghosting.

To illustrate the reliability of our proposed U-Net architecture, we included independently generated rat and mouse public datasets (Online dataset), including images acquired from different sites, scanners, and protocols. The presented results showed that U-Net produced stable and satisfactory results for both T2w RARE and T2*w EPI images. Although segmentation performance was not as robust in the mouse dataset, U-Net still reached significantly higher segmentation accuracy with averaged Dice > 0.85 for both T2w RARE and T2*w EPI compared to other methods, whereas the lowest averaged accuracy on all images was found in RATS (Dice = 0.82), PCNN (Dice = 0.79), and SHERM (Dice = 0.80) for mouse T2w RARE. This result highlights the reliable performance of the U-Net architecture for mouse brain MRI data.

There are several limitations of the U-Net architecture. First, deep learning is a data driven classification, so segmentation

accuracy highly relies on the training dataset. Indeed, we observed in **Supplementary Figures S1, S2** that manual segmentation accuracy is approximately the same as validation accuracy. Because we trained our U-Net algorithm by using only T2w RARE and T2*w EPI images in rats and mice, additional training and optimization will be needed to use our current U-Net architecture to skull-strip rodent brain images with different contrast (e.g., T1-weighted images). There are many challenges with conducting deep learning algorithm in multimodality datasets (i.e., heterogeneous sources, different levels of noise) (Ngiam et al., 2011; Baltrusaitis et al., 2019) as the features have to relate multiple data sources. Our future work will focus on developing rodent brain extraction tool specifically for T1w images. Second, deep learning methods require substantial amounts of manually labeled data (Verbraeken et al., 2020), and their performance can be affected by similarities between the training dataset and the unanalyzed dataset. The use of massive data augmentation is important in domains like biomedical segmentation, since the number of annotated samples is usually limited. More training datasets are needed to further improve our current U-Net architecture (e.g., including an additional mouse dataset with ground truth labels to improve our U-Net performance in mice). Third, our current U-Net architecture image patch limits the testing image to a matrix size of at least 128 × 128. Image resampling to a finer resolution is required if the image matrix size is smaller than 128 × 128. Fourth, whether 2D or 3D framework would yield better skull stripping or segmentation results remain an active topic of research (Baumgartner et al., 2018; Hänsch et al., 2018; Meine et al., 2018; Yu et al., 2019). Practically, each framework has its own advantages and disadvantages. For example, though 2D frameworks do not utilize information across slice direction and may only be suitable when slice resolution is coarse, they are also operationally efficient due to lower computational demands. Our results indeed support the feasibility of performing 2D U-Net framework in regular laptop CPU. The 3D framework, in contrast, preserves 3D context in training but suffers from inaccuracy when only limited number of slices is available. Finally, our future work will extend this study with more detailed classification of brain area labels so that automatic segmentation of brain nuclei using U-Net can be achieved.

CONCLUSION

The robustness of U-Net for delineating rodent brain boundaries are demonstrated in T2w RARE and T2*w EPI data acquired at multiple sites. Our proposed method demonstrated improved performance compared to current skull stripping methods, as determined using the qualitative metrics (Dice, Jaccard, PPV, SEN, and Hausdorff). We believe this tool will be useful to avoid parameter-selection bias and streamline pre-processing steps when analyzing rodent brain MRI data. Information about the CAMRI dataset used in this manuscript and our U-Net skull stripping tool can be found at <https://github.com/CAMRIatUNC/RodentMRISkullStripping>.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The CAMRI rats dataset is available at <https://openneuro.org/datasets/ds002870/versions/1.0.0> and mice dataset is available at <https://doi.org/10.18112/openneuro.ds002868.v1.0.0>. The U-Net skull stripping tool can be found at <https://github.com/CAMRIatUNC/RodentMRISkullStripping>. The names of the repository/repositories and accession number(s) can be found in the article/**Supplementary Material**.

ETHICS STATEMENT

Ethical review and approval was not required for the animal study because existing animal imaging database was used. No animal data acquired specifically for this project.

AUTHOR CONTRIBUTIONS

L-MH, DS, and Y-YS designed the study. L-MH and SW implemented U-Net algorithm for rodents. L-MH and PR validated the developed methods on various datasets. WB provided ground-truth brain masks. T-HC, SS, DC, LW, MB, and S-HL provided data and helped to edit the manuscript. S-HL managed data/software dissemination and helped to design

the study. L-MH and Y-YS wrote the manuscript. All authors contributed to the article and approved the submitted version.

FUNDING

This work was supported by the National Institute of Neurological Disorders and Stroke (R01NS091236), National Institute of Mental Health (RF1MH117053, R01MH111429, R41MH113252, and F32MH115439), National Institute on Alcohol Abuse and Alcoholism (P60AA011605, K01AA025383, and T32AA007573), and National Institute of Child Health and Human Development (P50HD103573).

ACKNOWLEDGMENTS

We thank Alicia Stevens at CAMRI for insightful discussion on this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnins.2020.568614/full#supplementary-material>

REFERENCES

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., et al. (2016). "TensorFlow: A system for large-scale machine learning," in *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016*, (California: USENIX Association).
- Albaugh, D. L., Salzwedel, A., Van Den Berge, N., Gao, W., Stuber, G. D., and Shih, Y. Y. I. (2016). Functional Magnetic Resonance Imaging of Electrical and Optogenetic Deep Brain Stimulation at the Rat Nucleus Accumbens. *Sci. Rep.* 6:31613. doi: 10.1038/srep31613
- Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., and Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging.* 6:014006. doi: 10.1117/1.jmi.6.1.014006
- Atay, S. M., Kroenke, C. D., Sabet, A., and Bayly, P. V. (2008). Measurement of the dynamic shear modulus of mouse brain tissue in vivo by magnetic resonance elastography. *J. Biomech. Eng.* 130:021013. doi: 10.1115/1.2899575
- Babalola, K. O., Patenaude, B., Aljabar, P., Schnabel, J., Kennedy, D., Crum, W., et al. (2009). An evaluation of four automatic methods of segmenting the subcortical structures in the brain. *Neuroimage* 47, 1435–1447. doi: 10.1016/j.neuroimage.2009.05.029
- Baltrusaitis, T., Ahuja, C., and Morency, L. P. (2019). Multimodal Machine Learning: A Survey and Taxonomy. *IEEE Trans. Pattern Anal. Mach. Intell.* 41, 423–443. doi: 10.1109/TPAMI.2018.2798607
- Baumgartner, C. F., Koch, L. M., Pollefeys, M., and Konukoglu, E. (2018). "An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation," in *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges. STACOM 2017. Lecture Notes in Computer Science*, ed. M. Pop (Cham: Springer).
- Bendazzoli, S., Brusini, I., Damberg, P., Smedby, Ö, Andersson, L., and Wang, C. (2019). Automatic rat brain segmentation from MRI using statistical shape models and random forest. *Image Proc.* 10949:1094920. doi: 10.1117/12.2512409
- Broadwater, M. A., Lee, S. H., Yu, Y., Zhu, H., Crews, F. T., Robinson, D. L., et al. (2018). Adolescent alcohol exposure decreases frontostriatal resting-state functional connectivity in adulthood. *Addict. Biol.* 23, 810–823. doi: 10.1111/adb.12530
- Chollet, F. (2015). *Keras Documentation*. Francisco: Keras.Io.
- Chou, N., Wu, J., Bai Bingren, J., Qiu, A., and Chuang, K. H. (2011). Robust automatic rodent brain extraction using 3-D pulse-coupled neural networks (PCNN). *IEEE Trans. Image Process.* 20, 2554–2564. doi: 10.1109/TIP.2011.2126587
- Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173. doi: 10.1006/cbmr.1996.0014
- Doshi, J., Erus, G., Ou, Y., Gaonkar, B., and Davatzikos, C. (2013). Multi-Atlas Skull-Stripping. *Acad. Radiol.* 20(12), 1566–76. doi: 10.1016/j.acra.2013.09.010
- Feo, R., and Giove, F. (2019). Towards an efficient segmentation of small rodents brain: A short critical review. *J. Neurosci. Methods* 323, 82–89. doi: 10.1016/j.jneumeth.2019.05.003
- Gaser, C., Schmidt, S., Metzler, M., Herrmann, K. H., Krumbein, I., Reichenbach, J. R., et al. (2012). Deformation-based brain morphometry in rats. *Neuroimage* 63, 47–53. doi: 10.1016/j.neuroimage.2012.06.066
- Grandjean, J., Canella, C., Anckaerts, C., Ayrancı, G., Bougacha, S., Bienert, T., et al. (2020). Common functional networks in the mouse brain revealed by multi-centre resting-state fMRI analysis. *Neuroimage* 205:116278. doi: 10.1016/j.neuroimage.2019.116278
- Grandjean, J., Corcoba, A., Kahn, M. C., Upton, A. L., Deneris, E. S., Seifritz, E., et al. (2019). A brain-wide functional map of the serotonergic responses to acute stress and fluoxetine. *Nat. Commun.* 10:350. doi: 10.1038/s41467-018-08256-w
- Hänsch, A., Schwier, M., Morgas, T., Klein, J., Hahn, H. K., Gass, T., et al. (2018). Comparison of different deep learning approaches for parotid gland segmentation from CT images. *Proc. SPIE* 10575:1057519. doi: 10.1117/12.2292962
- Havaei, M., Davy, A., Warde-Farley, D., Biard, A., Courville, A., Bengio, Y., et al. (2017). Brain tumor segmentation with Deep Neural Networks. *Med. Image Anal.* 35, 18–31. doi: 10.1016/j.media.2016.05.004

- Kingma, D. P., and Ba, J. L. (2015). "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, (Cornell: Cornell University).
- Kleesiek, J., Urban, G., Hubert, A., Schwarz, D., Maier-Hein, K., Bendszus, M., et al. (2016). Deep MRI brain extraction: A 3D convolutional neural network for skull stripping. *Neuroimage* 129, 460–469. doi: 10.1016/j.neuroimage.2016.01.024
- Konsman, J.-P. (2003). *The mouse brain in stereotaxic coordinates*. New York: Academic Press.
- Kuntimad, G., and Ranganath, H. S. (1999). Perfect image segmentation using pulse coupled neural networks. *IEEE Trans. Neural. Networks.* 10, 591–598. doi: 10.1109/72.761716
- Leung, K. K., Barnes, J., Modat, M., Ridgway, G. R., Bartlett, J. W., Fox, N. C., et al. (2011). Brain MAPS: An automated, accurate and robust brain extraction technique using a template library. *Neuroimage* 55, 1091–1108. doi: 10.1016/j.neuroimage.2010.12.067
- Liu, Y., Unsal, H. S., Tao, Y., and Zhang, N. (2020). Automatic Brain Extraction for Rodent MRI Images. *Neuroinformatics* 18, 395–406. doi: 10.1007/s12021-020-09453-z
- Lohmeier, J., Kaneko, T., Hamm, B., Makowski, M. R., and Okano, H. (2019). atlasBEX: Automated template-derived brain extraction in animal MRI. *Sci. Rep.* 9:12219. doi: 10.1038/s41598-019-4848948483
- Lu, H., Scholl, C. A., Zuo, Y., Demny, S., Rea, W., Stein, E. A., et al. (2010). Registering and analyzing rat fMRI data in the stereotaxic framework by exploiting intrinsic anatomical features. *Magn. Reson. Imaging.* 28, 146–152. doi: 10.1016/j.mri.2009.05.019
- Ma, Z., Perez, P., Ma, Z., Liu, Y., Hamilton, C., Liang, Z., et al. (2018). Functional atlas of the awake rat brain: A neuroimaging study of rat brain specialization and integration. *Neuroimage* 170, 95–112. doi: 10.1016/j.neuroimage.2016.07.007
- Madularu, D., Mathieu, A. P., Kumaragamage, C., Reynolds, L. M., Near, J., Flores, C., et al. (2017). A non-invasive restraining system for awake mouse imaging. *J. Neurosci. Methods.* 287, 53–57. doi: 10.1016/j.jneumeth.2017.06.008
- Mandino, F., Cerri, D. H., Garin, C. M., Straathof, M., van Tilborg, G. A. F., Chakravarty, M. M., et al. (2020). Animal Functional Magnetic Resonance Imaging: Trends and Path Toward Standardization. *Front. Neuroinform.* 13:78. doi: 10.3389/fninf.2019.00078
- Meine, H., Chlebus, G., Ghafourian, M., Endo, I., and Schenk, A. (2018). *Comparison of u-net-based convolutional neural networks for liver segmentation in ct*. arXiv Preprint. arXiv1810.04017.
- Milletari, F., Navab, N., and Ahmadi, S. A. (2016). "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proceedings - 2016 4th International Conference on 3D Vision, 3DV 2016*, (Cornell: Cornell University), doi: 10.1109/3DV.2016.79
- Ngiam, J., Khosla, A., Kim, M., Nam, J., Lee, H., and Ng, A. Y. (2011). "Multimodal deep learning," in *Proceedings of the 28th International Conference on Machine Learning, ICML 2011*, (Michigan: University of Michigan).
- Oguz, I., Zhang, H., Rumble, A., and Sonka, M. (2014). RATS: Rapid Automatic Tissue Segmentation in rodent brain MRI. *J. Neurosci. Methods.* 221, 175–182. doi: 10.1016/j.jneumeth.2013.09.021
- Paxinos, G., and Watson, C. (2014). *The Rat Brain in Stereotaxic Coordinates Seventh Edition*. Netherland: Elsevier.
- Pfefferbaum, A., Adalsteinsson, E., and Sullivan, E. V. (2004). In vivo structural imaging of the rat brain with a 3-T clinical human scanner. *J. Magn. Reson. Imaging.* 20, 779–785. doi: 10.1002/jmri.20181
- Ronneberger, O., Fischer, P., and Brox, T. (2015). "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, (Germany: University of Freiburg).
- Roy, S., Knutsen, A., Korotcov, A., Bosomtvi, A., Dardzinski, B., Butman, J. A., et al. (2018). "A deep learning framework for brain extraction in humans and animals with traumatic brain injury," in *Proceedings - International Symposium on Biomedical Imaging*, (New Jersey: IEEE).
- Sharief, A. A., Badea, A., Dale, A. M., and Johnson, G. A. (2008). Automated segmentation of the actively stained mouse brain using multi-spectral MR microscopy. *Neuroimage* 39, 136–145. doi: 10.1016/j.neuroimage.2007.08.028
- Shattuck, D. W., and Leahy, R. M. (2002). Brainsuite: An automated cortical surface identification tool. *Med. Image Anal.* 6, 129–142. doi: 10.1016/S1361-8415(02)00054-3
- Sirmpilatze, N., Baudewig, J., and Boretius, S. (2019). Temporal stability of fMRI in medetomidine-anesthetized rats. *Sci. Rep.* 9:16673. doi: 10.1038/s41598-019-53144-y
- Van Den Berge, N., Albaugh, D. L., Salzwedel, A., Vanhove, C., Van Holen, R., Gao, W., et al. (2017). Functional circuit mapping of striatal output nuclei using simultaneous deep brain stimulation and fMRI. *Neuroimage* 146, 1050–1061. doi: 10.1016/j.neuroimage.2016.10.049
- Verbraeken, J., Wolting, M., Katzy, J., Kloppenburg, J., Verbelen, T., and Rellermeyer, J. S. (2020). A Survey on Distributed Machine Learning. *ACM Comput. Surv.* 53:3377454. doi: 10.1145/3377454
- Wang, S., Nie, D., Qu, L., Shao, Y., Lian, J., Wang, Q., et al. (2020a). CT Male Pelvic Organ Segmentation via Hybrid Loss Network with Incomplete Annotation. *IEEE Trans. Med. Imaging.* 39, 2151–2162. doi: 10.1109/tmi.2020.2966389
- Wang, S., Wang, Q., Shao, Y., Qu, L., Lian, C., Lian, J., et al. (2020b). Iterative Label Denoising Network: Segmenting Male Pelvic Organs in CT from 3D Bounding Box Annotations. *IEEE Trans. Biomed. Eng.* 2020:99. doi: 10.1109/tbme.2020.2969608
- Yin, Y., Zhang, X., Williams, R., Wu, X., Anderson, D. D., and Sonka, M. (2010). LOGISMOS-layered optimal graph image segmentation of multiple objects and surfaces: Cartilage segmentation in the knee joint. *IEEE Trans. Med. Imaging.* 29, 2023–2037. doi: 10.1109/TMI.2010.2058861
- Yogananda, C. G. B., Wagner, B. C., Murugesan, G. K., Madhuranthakam, A., and Maldjian, J. A. (2019). "A deep learning pipeline for automatic skull stripping and brain segmentation," in *Proceedings - International Symposium on Biomedical Imaging*, (New Jersey: IEEE).
- Yu, Q., Xia, Y., Xie, L., Fishman, E. K., and Yuille, A. L. (2019). *Thickened 2D Networks for Efficient 3D Medical Image Segmentation*. New York: Cornell University. [Preprint]. arXiv1904.01150. Available online at: https://www.researchgate.net/publication/332168684_Thickened_2D_Networks_for_3D_Medical_Image_Segmentation (accessed January 05, 2020).
- Zhou, Z., Rahman Siddiquee, M. M., Tajbakhsh, N., and Liang, J. (2018). "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA 2018, ML-CDS 2018. Lecture Notes in Computer Science*, ed. D. Stoyanov (Cham: Springer).

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Hsu, Wang, Ranadive, Ban, Chao, Song, Cerri, Walton, Broadwater, Lee, Shen and Shih. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.