Published in final edited form as:

Proc SPIE Int Soc Opt Eng. 2021 February ; 11596: . doi:10.1117/12.2582264.

Accurate Estimation of Total Intracranial Volume in MRI using a Multi-tasked Image-to-Image Translation Network

Mallika Singh^a, Eleanor Pahl^b, Shangxian Wang^c, Aaron Carass^c, Junghoon Lee^d, Jerry L. Prince^{a,c}

^aDept. of Biomedical Engineering, The Johns Hopkins University, Baltimore, MD 21218

^bDept. of Aerospace Engineering, Embry-Riddle Aeronautical University (ERAU), Prescott, AZ 86301

^cDept. of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218

^dDept. of Radiation Oncology and Molecular Radiation Sciences, The Johns Hopkins School of Medicine, Baltimore, MD 21231

Abstract

Total intracranial volume (TIV) is the volume enclosed inside the cranium, inclusive of the meninges and the brain. TIV is extensively used to correct variations in inter-subject head size for the evaluation of neurodegenerative diseases. In this work, we present an automatic method to generate a TIV mask from MR images while synthesizing a CT image to be used in subsequent analysis. In addition, we propose an alternative way to obtain ground truth TIV masks using a semi-manual approach, which results in significant time savings. We train a conditional generative adversarial network (cGAN) using 2D MR slices to realize our tasks. The quantitative evaluation showed that the model was able to synthesize CT and generate TIV masks that closely approximate the reference images. This study also provides a comparison of the described method against skull stripping tools that output a mask enclosing the cranial volume, using MRI scan. In particular, highlighting the deficiencies in using such tools to approximate the volume using MRI scan.

Keywords

Human brain; MRI; Intracranial volume; Synthetic CT

1. INTRODUCTION

Total intracranial volume (TIV) comprises everything enclosed in the skull. This includes the gray matter, white matter, veins, and the meninges. TIV plays an important role in the study of neurodegenerative and neurodevelopmental disorders.¹ More importantly, it is used

as a normalization factor for brain volumes and thus improves the accuracy of analysis when evaluating brain structures.²⁻⁴

Magnetic resonance imaging (MRI) is one of the most widely used imaging modality to analyze the brain and brain structures due to its high-soft tissue contrast. However, manual estimation or delineation of TIV in MRI scans is time-consuming.^{5, 6} Hence, a TIV mask is often approximated from MRI scans using brain-extraction or skull-stripping tools that are publicly available in the form of software packages.⁶⁻⁹ Masks extracted from such tools are shown to be inaccurate,^{1, 10, 11} as they do not include the subarachnoid space or other non-brain TIV structures. Figure (1) shows the example where publicly available tools have been used to extract the volume mask. These masks have been superimposed on both the T1-w MRI image and the CT image. Since bone is not visible in conventional MRI scans, it is very challenging to estimate the cranium (skull-bone), which leads to errors when extracting TIV masks.

We present a deep learning framework that can estimate the cranium from MRI volumes while, simultaneously, segmenting the TIV masks. We consider a voxel-wise learning approach (based on Isola et al.¹²) that is modified to output our two tasks. Segmentations were created, for 20 MRI volumes from their corresponding co-registered CT volumes. TIV estimates from our deep-network, and various skull-stripping tools^{7-9, 13} were compared to the gold-standard manual delineation using surface distances and similarity measures. Synthetic CT (sCT) generated by the model were evaluated using image-based similarity and voxel-wise metrics.

2. DATASET

The data used in this work were obtained from the Radiation Oncology Department (Johns Hopkins School of Medicine, Baltimore, MD). The data comprised of twenty subjects, each consisting of T1-weighted MRI, T2-weighted MRI, and CT scans. Each subject's T1-w MR scan is bias-corrected and registered to the MNI space, prior to co-registering to the T2-w MR and CT scans. Scans in our dataset have been resampled to an isotropic resolution of 0.8mm x 0.8mm x 0.8mm.¹⁴

3. METHOD

3.1 Semi-manual delineation of TIV

Semi-manual TIV delineations are generated on the CT scans of all 20 subjects using ITK-SNAP¹⁵ and MIPAV¹⁶ software packages. The delineations are produced in a two-step approach that makes use of the Active Contour Segmentation method (included in ITK-SNAP toolbox) to develop an initial TIV segmentation mask that is manually corrected to ensure there is no bleed-through in the sinuses. Guidance from brain atlases,^{17, 18} around complicated regions of the skull consisting of microstructures and sinuses, helped determine which voxels were to be included in the mask. McRae's Line, which defines the base of the skull, was the defined lower bound of our TIV delineations (Fig. 2d).

All delineations were performed by a single rater and reviewed by two observers. The recorded time taken to generate the TIV mask was approximately 20 min/scan for our two-step delineation approach, which is a significant improvement over the 1-2 hours using a traditional manual delineation. No time limitation was imposed when delineating the TIV masks.

3.2 Automated TIV estimation and CT synthesis

The core component of our method is a synthesis-segmentation network which was developed based on the image-to-image translation network, or pix2pix concept, introduced by Isola et al.¹² Pix2pix has been widely used in the synthesis of images and shows promising segmentation predictions.^{19, 20} Image-to-image translation networks are a conditional GAN consisting of a generator-discriminator pair, both conditioned on the target image. The architecture described in 12, uses a UNet²¹ encoder-decoder generator, with the tanh activation function in the output layer to ensure the generated image is bounded in the range $[-1, 1]$. The method uses a PatchGAN discriminator that estimates how realistic the generated output is when compared to the target.

Within this work, the network is modified to a single generator-dual discriminator model (Figure 3). The output of the decoder branch in the generator network is split to produce two different outputs (Figure 4). The encoder takes in input MR slices concatenated along the channel axis, i.e., the size of the input image is $2 \times 256 \times 256$ with 2 representing the number of channels. The encoder is made up of 8 contracting blocks, with each contracting block consists of a convolution layer, Batch norm layer, and Leaky ReLU activation layer. While the decoder consists of 9 expanding blocks with 7 common blocks and one expanding block for each head to generate single channel output images of size $1 \times 256 \times 256$. Each expanding block is made of up a transpose convolution layer, batch normalization layer, and a ReLU activation layer. Dropout has been added to the first three expanding blocks.

4. RESULTS

The 20 data sets were separated into a training set containing MR, CT volumes, and the annotations of 14 subjects and a separate test and validation set containing volumes corresponding to 4 and 2 subjects, respectively. Each MR or CT volume is comprised of 241 axial 2D image slices. These were padded and cropped to 256×256 pixel images, with intensity values distributed in $[-1000, 1500]$ HU for CT. Intensities outside the accepted range are set to the nearest boundary value. MR scans white-matter normalized. All scans were normalized to be within $[-1, 1]$ intensity range, to facilitate the training. Input to the network consists of multi-modal MRI images (T1-w and T2-w), while the source-target consists of the corresponding CT and the delineated TIV mask. To increase the data size, we used data augmentation techniques such as random flipping, random rotation (limited up to 30 degrees). The model was trained for 2000 epochs using Adam optimizer. The majority of the training scheme remained similar to the one described by Isola et al.¹²

4.1 CT Synthesis:

Figure 5 shows an example of the input MR slices, the synthesized CT image generated by the model, the corresponding source target CT image, and the absolute difference image. The difference image shows the absolute error between the target and generated CT. The model has learned to transform intensities of soft-brain tissues but not those surrounding the bone structures, i.e., differences are more pronounced near the orbit (eye socket), sinuses, and base of the skull (Figure 6, case 2). The difference image also shows the intensity difference along the contour of the head and air.

The similarity between the target CT and generated CT scans (for four test subjects), was assessed using three comparison metrics: the mean absolute error (MAE) metric, computed in HU (Hounsfield Units); the structural similarity index measure (SSIM); and peak-signal-to-noise ratio (PSNR) [Table 1]. Quantitative results show that the described model outperforms the baseline pix2pix network in terms of the MAE, SSIM, and peak-SNR.

4.2 Segmentation:

To evaluate the presented methodology, we compared the results of the segmentation output with the existing methods used to estimate the intracranial volume (Figure 7). Skull-stripping methods such as MONSTR,⁸ FSL-brain extraction tool (BET),⁷ HD-BET,⁹ and ROBust Brain Extraction tool (ROBEX),¹³ are used to obtain a binary mask to approximate the TIV. Accuracy of TIV extraction is evaluated by assessing the overlap between the ground truth and estimated masks, and surface distance maps were computed to evaluate surface dissimilarities (Table 2).

Quantitative evaluations show that the masks generated by the described method performed better in terms of estimating the surface volume, with an average surface difference of 1.980 ± 0.661 . The similarity between the extracted volume and the annotated volume, show that our method seemed to be more reliable compared to the other methods.

5. CONCLUSION

We present a multi-task deep network approach for synthesizing CT images and segmenting the intracranial volume from multi-modal MRI images. The network extends the well-established pix2pix network to perform supervised synthesis and segmentation tasks simultaneously. In the application of synthesis, it is evident that the generated CT images fared well in terms of visual quality and performance metrics. However, we notice that synthesis towards the base of the skull is lacking in comparison to supervised²² and unsupervised methods.²³

In the application of volume mask segmentation, the generated binary masks are visually similar when validated against existing volumes extracted using skull-stripping methods and the ground-truth. On further analysis, we found our method to more accurately approximate the TIV than the ones described in comparison to our ground-truth delineations. However, we are aware that since the network is dependent on the semi-manual delineations that were used as the ground truth, it might be biased. A recommended approach for future work is to

conduct an intra and inter-delineator reliability study. Future work will focus on employing a 3D CNN, which is expected to bring further improvements, and working with subjects from a different cohort, that should help us address issues with domain shift.²⁴

REFERENCES

- [1]. Khalili N, Turk E, Benders M, Moeskops P, Claessens N, Heus R, Franx A, Wagenaar N, Breur H, Viergever M, and Isgum I, "Automatic extraction of the intracranial volume in fetal and neonatal mr scans using convolutional neural networks," *NeuroImage: Clinical* 24 (11 2019).
- [2]. Sargolzaei S, Sargolzaei A, Cabrerizo M, Chen G, Goryawala M, Noei S, Zhou Q, Duara R, Barker W, and Adjouadi M, "A practical guideline for intracranial volume estimation in patients with alzheimer's disease," *BMC bioinformatics* 16 Suppl 7, S8 (4 2015).
- [3]. Goto M, Hagiwara A, Kato A, Fujita S, Hori M, Kamagata K, Sugano H, Arai H, Aoki S, Abe O, Sakamoto H, Sakano Y, Kyogoku S, and Daida H, "Estimation of intracranial volume: A comparative study between synthetic mri and fsl-brain extraction tool (bet)2," *Journal of Clinical Neuroscience* 79, 178–182 (8 2020). [PubMed: 33070892]
- [4]. Glaister J, Carass A, NessAiver T, Stough JV, Saidha S, Calabresi PA, and Prince J, "Thalamus segmentation using multi-modal feature classification: Validation and pilot study of an age-matched cohort," *NeuroImage* 158, 430–440 (2 2017). [PubMed: 28669906]
- [5]. Klasson N, Olsson E, Rudemo M, Eckerström C, Malmgren H, and Wallin A, "Valid and efficient manual estimates of intracranial volume from magnetic resonance images," *BMC Medical Imaging* 15 (12 2015).
- [6]. Keihaninejad S, Heckemann RA, Fagiolo G, Symms MR, Hajnal JV, Hammers A, and The Alzheimer's Disease Neuroimaging Initiative, "A robust method to estimate the intracranial volume across MRI field strengths (1.5T and 3T)," *NeuroImage* 50(4), 1427–1437 (2010). [PubMed: 20114082]
- [7]. Smith S, "Fast robust automated brain extraction," *Human Brain Mapping* 17(3), 143–155 (11 2002). [PubMed: 12391568]
- [8]. Roy S, Butman JA, Pham DL, and Alzheimers Disease Neuroimaging Initiative, "Robust skull stripping using multiple mr image contrasts insensitive to pathology," *NeuroImage* 146, 132–147 (2 2017). [PubMed: 27864083]
- [9]. Isensee F, S. M, Pflueger I BG, Bonekamp D, Neuberger U, Wick A, Schlemmer H, Heiland S, Wick W, Bendszus M, Maier-Hein K, and Kickingereder P, "Automated brain extraction of multisequence mri using artificial neural networks," *Human Brain Mapping* 40(17), 4952–4964 (12 2019). [PubMed: 31403237]
- [10]. Nordenskjöld R, Malmberg F, Larsson E-M, Simmons A, Brooks SJ, Lind L, Ahlström H, Johansson L, and Joel K, "Intracranial volume estimated with commonly used methods could introduce bias in studies including brain volume measurements," *NeuroImage* 83 (12 2013).
- [11]. Voevodskaya O, Simmons A, Nordenskjöld R, Kullberg J, Ahlström H, Lind L, Wahlund L-O, Larsson E-M, Westman E, and Alzheimer's Disease Neuroimaging Initiative, "The effects of intracranial volume adjustment approaches on multiple regional MRI volumes in healthy aging and Alzheimer's disease," *Frontiers in Aging Neuroscience* 6 (2014).
- [12]. Isola P, Zhu J, Zhou T, and Efros AA, "Image-to-image translation with conditional adversarial networks," 5967–5976 (6 2017).
- [13]. Iglesias J, Liu C, Thompson P, and Tu Z, "Robust brain extraction across datasets and comparison with publicly available methods," *IEEE Transactions on Medical Imaging* 30(9) (2011).
- [14]. Zhao C, Shao M, Carass A, Li H, Dewey BE, Ellingsen LM, Woo J, Guttman MA, Blitz AM, Stone M, Calabresi PA, Halperin H, and Prince JL, "Applications of a deep learning method for anti-aliasing and super-resolution in mri," *Magnetic Resonance Imaging* 64, 132–141 (2019). [PubMed: 31247254]
- [15]. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, and Gerig G, "User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability," *NeuroImage* 31 (2006).

- [16]. McAuliffe MJ, Lalonde FM, McGarry D, Gandler W, Csaky K, and Trus BL, "Medical image processing, analysis and visualization in clinical research," in [Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001], 381–386 (2001).
- [17]. Peris-Celda M, Martínez-Soriano F, and Rhoton AL, "Rhoton's Atlas of Head, Neck, and Brain: 2D and 3D Images," (2018).
- [18]. Moeller T and Reif E, [Pocket Atlas of Sectional Anatomy, Computed Tomography and Magnetic Resonance Imaging, Vol. 2: Thorax, Heart, Abdomen, and Pelvis], Thieme, New York (2007 (third edition)).
- [19]. Eslami M, Tabarestani S, Albarqouni S, Adeli E, Navab N, and Adjouadi M, "Image-to-images translation for multi-task organ segmentation and bone suppression in chest x-ray radiography," IEEE Transactions on Medical Imaging 39(7) (2020).
- [20]. Liu F, Cai J, Huo Y, Cheng C-T, Raju A, Jin D, Xiao J, Yuille A, Lu L, Liao C, and Harrison AP, "Jssr: A joint synthesis, segmentation, and registration system for 3d multi-modal image alignment of large-scale pathological ct scans," (2020).
- [21]. Ronneberger O, Fischer P, and Brox T, "U-net: Convolutional networks for biomedical image segmentation," Medical Image Computing and Computer-Assisted Intervention – MICCAI (2015).
- [22]. Hiasa Y, Otake Y, Takao M, Matsuoka T, Takashima K, Carass A, Prince J, Sugano N, and Sato Y, "Cross-modality image synthesis from unpaired data using cyclegan: Effects of gradient consistency loss and training data size," Simulation and Synthesis in Medical Imaging held in conjunction with the 21st International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI), Granada, Spain, 9 16–20 , 31–41 (2018).
- [23]. Yang H, Sun J, Carass A, Zhao C, Lee J, Xu Z, and Prince JL, "Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan," IEEE Trans Med Imaging. 39(12), 4249–426 (2020). [PubMed: 32780700]
- [24]. Murez Z, Kolouri S, Kriegman D, Ramamoorthi R, and Kim K, "Image to image translation for domain adaptation," in [Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)], (6 2018).

Figure 1:
Extracted masks super-imposed on T1-weighted MR and CT images for visualization purposes. Arrows indicate the presence of bone in both the MR and CT images, **(a)** is the MONSTR mask and **(b)** is the BET mask.

Figure 2:

Semi-manual delineation: (a)-(c) indicate structures, such as the internal carotid artery (a) and the sella turcica (b,c), to be included in the TIV masks. (d) The lower bound of the TIV is defined by McRae's line, indicated by the line.

Figure 3:
Conceptual block diagram.

Figure 5:

From left to right: T1-w MR image, T2-w MR image, synthesized CT image, source target CT image, and absolute error between real and synthesized CT image.

Figure 6:
Real (b,d) and generated synthetic CT (a,c) images

Figure 7:

Total Intracranial Volume (TIV) masks: From left to right, (a) the semi-manually delineated mask, masks extracted from (b) MONSTR, (c) BET, (d) ROBEX, (e) HD-BET and (f) our network generated mask, super-imposed on the native CT images for better visualization. Highlighted are the regions of over and underestimation.

Table 1:

Comparison metrics between synthesized and real CT images pertaining to the four test subjects: Mean absolute error (MAE) values in HU, Structural Similarity Index Measure (SSIM) and peak-signal-to-noise ratio (PSNR).

	CT Synthesis Method	
	Pix2Pix	Our Method
MAE (HU)	76.46 $\pm$ 13.15	53.14 $\pm$ 15.77
SSIM	0.799 $\pm$ 0.06	0.853 $\pm$ 0.05
PSNR	25.26 $\pm$ 0.82	27.01 $\pm$ 0.76

Table 2:

95th percentile Hausdorff Distance (HD95), Dice similarity coefficient (DSC), and the Jaccard similarity (Jacc) among various skull-stripping tool and TIV measurement methods.

	Generated Masks				
	MONSTR	FSL-BET	ROBEX	HD-BET	Our Method
HD95 (mm)	3.932 ± 0.225	6.197 ± 2.113	6.472 ± 0.383	5.242 ± 0.602	1.980 ± 0.661
Dice	0.974 ± 0.003	0.952 ± 0.014	0.954 ± 0.005	0.962 ± 0.006	0.984 ± 0.006
Jacc	0.950 ± 0.006	0.909 ± 0.025	0.913 ± 0.009	0.927 ± 0.011	0.969 ± 0.011