



Published in final edited form as:

Neuroinformatics. 2022 July ; 20(3): 651–664. doi:10.1007/s12021-021-09544-5.

Convolutional Neural Network based frameworks for fast automatic segmentation of thalamic nuclei from native and synthesized contrast structural MRI

Lavanya Umapathy^{1,2}, Mahesh Bharath Keerthivasan^{2,5}, Natalie M. Zahr³, Ali Bilgin^{1,2,4}, Manojkumar Saranathan^{1,2,4}

¹Department of Electrical and Computer Engineering, University of Arizona, Tucson, AZ, United States

²Department of Medical Imaging, University of Arizona, Tucson, AZ, United States

³Department of Psychiatry & Behavioral Sciences, Stanford University, Menlo Park, CA, United States

⁴Department of Biomedical Engineering, University of Arizona, Tucson, AZ, United States

⁵Siemens Medical Solutions USA, Tucson, AZ, United States

Abstract

Thalamic nuclei have been implicated in several neurological diseases. Thalamic nuclei parcellation from structural MRI is challenging due to poor intra-thalamic nuclear contrast while methods based on diffusion and functional MRI are affected by limited spatial resolution and image distortion. Existing multi-atlas based techniques are often computationally intensive and time-consuming. In this work, we propose a 3D Convolutional Neural Network (CNN) based framework for thalamic nuclei parcellation using T1-weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) images. Transformation of images to an efficient representation has been proposed to improve the performance of subsequent classification tasks especially when working with limited labeled data. We investigate this by transforming the MPRAGE images to White-Matter-nulled MPRAGE (WMn-MPRAGE) contrast, previously shown to exhibit good intra-thalamic nuclear contrast, prior to the segmentation step. We trained two 3D segmentation frameworks using MPRAGE images (n=35 subjects): a) a native contrast segmentation (NCS) on MPRAGE images and b) a synthesized contrast segmentation (SCS) where synthesized WMn-MPRAGE representation generated by a contrast synthesis CNN were used. Thalamic nuclei

Corresponding Author: Manojkumar Saranathan, PhD, Associate Professor, Department of Medical Imaging, University of Arizona, Tucson, AZ 85724, anagrammian@gmail.com, Phone: (520) 626-6531.

Conflict of Interest

The authors have no conflicts of interest, relevant to this manuscript, to disclose. Additional disclosures: MBK (spouse of LU) is an employee of Siemens Medical Solutions, USA.

Availability of Data and Materials

The datasets utilized for the training and testing of the CNNs are currently not publicly available due to IRB data sharing restrictions. On acceptance, a smaller subset of anonymized data would be made available with the GitHub repository for testing purposes. Access to datasets will be available from the corresponding author on reasonable request.

Code Availability

The code used in this work is available on GitHub (<https://github.com/lunastra26/thalamic-nuclei-segmentation>).

labels were generated using THOMAS, a multi-atlas segmentation technique proposed for WMn-MPRAGE images. The segmentation accuracy and clinical utility were evaluated on a healthy cohort (n=12) and a cohort (n=45) comprising of healthy subjects and patients with alcohol use disorder (AUD), respectively.

Both the segmentation CNNs yielded comparable performances on most thalamic nuclei with Dice scores greater than 0.84 for larger nuclei and at least 0.7 for smaller nuclei. However, for some nuclei, the SCS CNN yielded significant improvements in Dice scores (medial geniculate nucleus, $P=.003$, centromedian nucleus, $P=.01$) and percent volume difference (ventral anterior, $P=.001$, ventral posterior lateral, $P=.01$) over NCS. In the AUD cohort, the SCS CNN demonstrated a significant atrophy in ventral lateral posterior nucleus in AUD patients compared to healthy age-matched controls ($P=0.01$), agreeing with previous studies on thalamic atrophy in alcoholism, whereas the NCS CNN showed spurious atrophy of the ventral posterior lateral nucleus.

CNN-based segmentation of thalamic nuclei provides a fast and automated technique for thalamic nuclei prediction in MPRAGE images. The transformation of images to an efficient representation, such as WMn-MPRAGE, can provide further improvements in segmentation performance.

Keywords

Representational learning; Contrast synthesis; White-matter-nulled MPRAGE; Thalamic-nuclei segmentation; 3D Convolutional Neural Networks; Alcohol Use Disorder

INTRODUCTION

The thalamus, with its histologically and functionally distinct nuclei, is a deep brain relay organ involved in higher level neurological functions including sleep, arousal, and consciousness. Individual thalamic nuclei have been shown to be differentially involved in multiple neurological diseases including mediodorsal (MD) and pulvinar (Pul) nuclei for schizophrenia (Byne et al. 2001), interlaminar nuclei for Parkinson's Disease (Henderson et al. 2000), limbic and anterior thalamic nuclei for Alzheimer's disease (Braak and Braak 1991; Aggleton et al. 2016; Bernstein et al. 2021), and anterioventral (AV), Pul, and habenular (Hb) nuclei for multiple sclerosis (Planche et al. 2019). Segmentation and analysis of individual thalamic nuclei, in addition to the thalamus, can not only provide more insight in understanding the relationship between thalamic volumes and neurological diseases but also better elucidate the role of thalamic nuclei in healthy and aging brains (Fama and Sullivan 2015).

Many automated methods for parcellation of thalamic nuclei reported in the literature are based on diffusion-MRI based techniques (Wiegell et al. 2003; Battistella et al. 2017; Behrens et al. 2003). Diffusion MRI inherently suffers from poor spatial resolution of the underlying echo-planar imaging acquisition. This along with the isotropic nature of diffusion within the thalamus, a predominantly gray matter region, has limited diffusion-MRI based segmentation to either larger nuclei or nuclei clusters. Recently, Iglehart et al (2020) presented a systematic comparison of thalamic nuclei segmentation techniques based on structural, diffusion, and resting state functional MR acquisitions. In this work,

the authors showed that parcellations on structural MR images were more accurate in the delineation of smaller thalamic structures such as AV, Hb, and medial geniculate nucleus (MGN) compared to Diffusion Tensor Imaging (DTI). For thalamic nuclei segmentation on structural T1-weighted scans such as Magnetization Prepared Rapid Gradient Echo (MPRAGE), Iglesias et al (2018) developed a probabilistic thalamus atlas from ex-vivo histology and in-vivo MPRAGE and used Bayesian inference to segment thalamic nuclei on in vivo MR images. Available as part of the neuroimaging package Freesurfer (<https://surfer.nmr.mgh.harvard.edu>), this segmentation technique is computationally intensive and time-consuming. Liu et al (2020) presented a hierarchical statistical shape model generated from histological-like atlases derived from manual delineation of thalamic nuclei in a series of 7T MR images acquired on 9 healthy subjects. Using shape-based models for thalamic nuclei segmentation, the authors reported average Dice scores of 0.53 to 0.90 for all nuclei with leave-one-out validation on healthy subjects. The segmentation accuracy relies on the shape-based multi-atlas technique but does not make use of intra-thalamic intensity information. Although routine clinical studies include MPRAGE scans, these images suffer from poor contrast (Saranathan et al. 2015) between the thalamus and its surrounding tissues as well as poor intra-thalamic nuclear contrast, making manual or automated thalamic nuclei segmentation challenging.

Recent works have demonstrated that a white-matter-nulled variant of the MPRAGE sequence, hereafter referred to as WMn-MPRAGE, provides improved image contrast to clearly identify thalamus boundaries for segmentation (Tourdias et al. 2014), and maximize intra-thalamic nuclear contrast at 7T (Saranathan et al. 2015). A multi-atlas technique called Thalamus Optimized Multi Atlas Segmentation (THOMAS) (Su et al. 2019) was proposed to segment thalamic nuclei on WMn-MPRAGE images at 7T. THOMAS was shown to perform well when validated against manual annotations, with Dice scores of at least 0.85 for the larger thalamic nuclei and at least 0.7 for smaller structures. Although WMn-MPRAGE provides excellent intra-thalamic nuclei contrast, this sequence is not part of conventional clinical imaging protocols and its inclusion will increase the overall scan time by 7-10 minutes. Moreover, numerous online data repositories that are used to study brain structure and metabolism like Alzheimer's Disease Neuroimaging Initiative (ADNI) or Open Access Series of Imaging Studies (OASIS) contain only MPRAGE images.

Supervised deep learning models, especially Convolutional Neural Networks (CNNs), have been shown to provide state-of-the-art performance in medical image segmentation tasks in the presence of large, labeled datasets (Litjens et al. 2017), with faster prediction times and performance comparable to human observer variability. When working with limited labeled data, as is often the case with medical images, transformation of data to an efficient representation can leverage this limited availability and make the subsequent classification task easier (Ian Goodfellow and Yoshua Bengio and Aaron Courville 2016). The choice of this representation often depends upon the classification task under investigation and could be as simple as mapping the original image contrast to a new contrast where prior domain knowledge has demonstrated possibility for improved classification performance. The use of CNNs for synthesizing contrasts has been explored in various MRI applications including inter-modality and intra-modality synthesis (Wang et al. 2021). These techniques aim to improve pathology detection without the overhead of additional scan time associated with

acquiring a new contrast. For example, synthesis of FLAIR (Wei et al. 2019; Hagiwara et al. 2019) and double inversion recovery contrasts (Finck et al. 2020) for multiple sclerosis detection, synthesis of MP2RAGE contrast from MPRAGE images for automatic lesion segmentation (La Rosa et al. 2021) and synthesis of magnetic resonance angiography from conventional T1 and T2-weighted MRI (Olut et al. 2018). Moreover, these synthesized images have also been shown to aid subsequent post-processing tasks such as tumor segmentation (Conte et al. 2021; La Rosa et al. 2021) and image registration (Yang et al. 2020).

In this work, we propose a 3D deep learning based multi-scale CNN framework for thalamic nuclei segmentation using MPRAGE images. We refer to this segmentation CNN framework using MPRAGE image contrast as native contrast segmentation (NCS). We also explore the synthesis of WMn-MPRAGE contrast as an efficient representation of MPRAGE images and introduce a 3D CNN-based synthesis model prior to the segmentation CNN to learn this representational mapping from MPRAGE to WMn-MPRAGE contrast. We investigate if this representational learning can further improve thalamic nuclear conspicuity and, in turn, segmentation accuracy. We refer to the generated representation as synthesized WMn-MPRAGE and the overall segmentation framework as synthesized contrast segmentation (SCS).

MATERIALS AND METHODS

CNN Architecture Overview

Figure 1 shows an overview of the proposed thalamic nuclei segmentation frameworks. Fig. 1a shows the 3D Native Contrast Segmentation (NCS) CNN that predicts the whole thalamus and individual thalamic nuclei on MPRAGE images. Fig. 1b shows the Synthesized Contrast Segmentation (SCS) CNN where a 3D contrast synthesis CNN to generate synthesized WMn-MPRAGE representation from MPRAGE images is introduced prior to the segmentation CNN.

The multi-scale thalamic nuclei segmentation CNN (Fig. 2a) has a U-NET (Ronneberger, Fischer, and Brox 2015) like encoder-decoder framework that processes features at multiple scales. At each resolution level, a series of convolution (3x3x3), batch normalization, and rectified linear unit layers extract features followed by a pooling layer for downsampling. The skip connections from each resolution level are concatenated to the corresponding resolution level along the upsampling path. The segmentation CNN uses MPRAGE images as input and predicts both the thalamus and individual thalamic nuclei in one pass through the model. Multiresolution feature maps are first used to predict the thalamus label. These feature maps, concatenated with the predicted thalamus label from the previous step, are then used to localize individual thalamic nuclei. Whole thalamus and thalamic nuclei labels generated from WMn-MPRAGE images using THOMAS are used as surrogate for manual segmentation ground truth for training the segmentation CNN. Dice loss is used for the thalamus and multi-label Dice loss is used for thalamic nuclei. The total segmentation loss is the sum of errors in the prediction of the thalamus using Dice loss and its individual nuclei using multi-label Dice loss. If C refers to the total number of nuclei labels, G_j and P_j refer

to the ground truth annotations and predictions for the i^{th} label, respectively, the multi-label Dice loss used in this work is defined as follows:

$$L_{\text{multi-label Dice}} = \frac{1}{C} \left(1 - \sum_i \frac{2 |G_i \cap P_i|}{|G_i| + |P_i|} \right)$$

The architecture for the multi-scale contrast synthesis CNN also uses an encoder-decoder architecture similar to the segmentation CNN (Fig. 2b) except for the loss formulation. The contrast synthesis CNN is trained using MPRAGE images as input and co-registered WMn-MPRAGE images as ground truth. Here, along with an intensity similarity-based loss function, we also use the perceptual loss measured using a VGG16 model (Simonyan and Zisserman 2015) pre-trained on ImageNet, similar to some other works on contrast synthesis in MR images (Dar et al. 2019). If W and W_{syn} are the ground truth WMn-MPRAGE image and the WMn-MPRAGE image synthesized from the corresponding MPRAGE image, respectively, the total synthesis loss function is given as:

$$L\{W, W_{syn}\} = \|W - W_{syn}\|_1 + \|F(W) - F(W_{syn})\|_2$$

The first term represents the intensity similarity-based loss defined as the L_1 norm-error between the ground truth and the synthesized images. The second term, corresponding to the L_2 norm error of the perceptual loss, ensures visual similarity of the synthesized images and the reference images. The function $F(\cdot)$ extracts feature maps from the third rectified linear unit layer of the ImageNet pre-trained VGG16 model to compare features from ground truth and synthesized images.

Study Cohort

For generalizability, images acquired on two different scanners were pooled for training. This comprised of data from 18 healthy individuals acquired on a 3T MAGNETOM Prisma (Siemens Medical Solutions USA, Inc., Malvern, PA, USA) and 17 healthy individuals acquired on a 3T Signa (General Electric Healthcare, Waukesha, WI) MR scanner, each subject being scanned with both MPRAGE and WMn-MPRAGE protocols. Of these 35 individuals, 33 subjects were retained for training, 2 were used for validation. Note that the training and validation cohorts consisted entirely of healthy individuals with no known pathologies.

Two independent test cohorts were used to demonstrate generalizability. The performance of the thalamic nuclei segmentation CNN trained directly on MPRAGE images (i.e. NCS) with those using synthesized WMn-MPRAGE images (i.e. SCS) is first evaluated on a cohort of healthy subjects ($n=12$). This cohort is referred to as test cohort 1. For clinical validation, the CNNs proposed in this work are used on MPRAGE images from an independent cohort of subjects ($n=45$) with alcohol use disorder (AUD) ($n=22$) and gender- and age-matched controls ($n=23$) to study the effect of AUD on thalamic nuclear atrophy. This second test cohort is referred to as the AUD cohort. The WMn-MPRAGE images from this cohort were previously used to characterize atrophy in thalamic nuclei volumes in alcoholism

using THOMAS (Zahr et al. 2020). The image acquisition for the testing cohorts also consisted of MPRAGE and WMn-MPRAGE protocols. All imaging data was acquired after prior informed consent and adhering to institutional review board guidelines. Additional information on MR image acquisition protocols for the subjects used in this study are available in the Supplementary Table 1.

CNN Data Generation

The ground truth labels for the thalamus and 11 thalamic nuclei were generated on WMn-MPRAGE images using THOMAS (Su et al. 2019), a multi-atlas segmentation technique. THOMAS uses 20 anatomical priors with manual delineations by a radiologist, guided by the Morel atlas, to segment the thalamus into 11 nuclei and the mammillothalamic tract in 15 minutes or less.

The isotropic resolution MPRAGE images and the corresponding WMn-MPRAGE images were co-registered using an affine transformation to train the segmentation CNNs. The co-registered images were brain extracted using FSL's brain extraction toolbox (Smith 2002), N4-bias corrected (Tustison et al. 2010), and intensity normalized using histogram-based contrast stretching to scale intensities in the range of [0, 1]. Contrast stretching was performed to clip intensities below the 1st percentile and above the 99th percentile to avoid outlier-based clipping that may affect traditional min-max based intensity normalization.

The segmentation CNNs were trained using an image-based scheme. From each subject, training data for the segmentation CNN were generated by cropping and extracting 2.5D (192x192x5) images using a sliding window. Data augmentation for the segmentation CNN training included scaling and shearing transformations. The contrast synthesis CNN was trained using a patch-based scheme. The co-registered image pairs (MPRAGE and WMn-MPRAGE images) were used to extract overlapping 2.5D patches (64x64x5) using a sliding window over the entire brain. The patches were extracted only from regions within the brain mask for each subject. Data augmentation for the synthesis CNN included random in-plane and through-plane rotations in addition to scaling and shearing transformations. Augmented images were generated on every training image batch by randomly selecting a subset of data augmentation schemes described previously during run time.

CNN Experiments and Implementation

To understand how segmentation CNN performance changes with loss function, we also trained the NCS CNN with weighted categorical cross entropy loss (WCCE). Here, a binary cross-entropy loss was used for thalamus and WCCE was used for thalamic nuclei. The WCCE loss function is defined as:

$$L_{wcce} = - \sum_n \sum_i^C w_i g_n \log p_n$$

Here, C is the number of labels, w_j is the weight associated with the i^{th} label, g_n and p_n refer to the n^{th} pixel of the ground truth (G) and prediction (P), respectively.

The training parameters for the thalamic nuclei segmentation CNN are: loss function for thalamus = Dice, loss function for thalamic nuclei = multi-label Dice, epochs = 50, batch size = 10, optimizer = Adam, learning rate = 0.001, and decay = 0.1. The training parameters for the contrast synthesis CNN are: Loss function for image similarity = L1, perceptual loss = L2, epochs = 50, batch size = 10, optimizer = Adam, learning rate = 0.001, decay = 0.1. Experiments were implemented in Python using Keras (<https://keras.io>) with Tensorflow (<https://www.tensorflow.org>) backend on a Linux system, with Titan P100 (NVIDIA) GPU. Architecture details and code used in this work are available at the following URL: <https://github.com/lunastra26/thalamic-nuclei-segmentation>.

Evaluation of CNN Performance

The whole thalamus and individual thalamic nuclei labels were predicted on MPRAGE images using the two segmentation frameworks. Performance of the segmentation CNN was evaluated using Dice score and percent volume difference (VD). If G and P refer to the ground truth THOMAS labels and predictions, respectively, then, $\text{Dice} = \frac{2 |G \cap P|}{|G| + |P|}$ and $\text{VD}(\%) = \frac{|V_G - V_P|}{V_G} \times 100$. Here, V_G corresponds to the volume of mask G . The performance of the contrast synthesis CNN was evaluated using root mean squared error (RMSE), structural similarity (SSIM), and peak signal-to-noise (PSNR) ratio. The definitions for these evaluation metrics are presented in the Supplementary material.

Statistical Analysis

A paired two-sided t-test ($\alpha = 0.05$) was used to compare if there were any significant differences in the performance of the segmentation CNNs trained on MPRAGE images compared to their WMn-MPRAGE representation. A Bland Altman analysis was performed to assess agreement between ground truth and volumes predicted by NCS and SCS CNNs. Pearson's correlation coefficient, coefficient of variation (CV) and percent reproducibility coefficient (RPC) statistics were also computed.

A recent work by Zahr et al (2020) used WMn-MPRAGE images and THOMAS segmentation to detect significant atrophy in the ventral lateral posterior (VLP) nucleus on a cohort comprising of patients with alcohol use disorder (AUD) and age-matched controls. We used the AUD cohort from this work to characterize atrophy in thalamic nuclei volumes in alcoholism, but using MPRAGE images instead of WMn-MPRAGE. The 11 thalamic nuclei were combined into four groups similar to Zahr et al (2020) : anterior, lateral, posterior, and medial. The nuclei breakdown for these groups is shown in Table 1. ANCOVA analysis was first performed on the predicted volumes from the NCS and SCS CNNs to understand the relationship between volumes of the nuclei groups and diagnosis (controls vs alcoholics), after accounting for age and intercranial volume (ICV) estimated using Freesurfer segmentation (Buckner et al. 2004). If a significant relationship was observed within any of these groups, ANCOVA analysis was repeated for individual nuclei within that group to further understand relationship between thalamic nuclei volumes and diagnosis.

RESULTS

The training of the segmentation and synthesis CNNs took approximately 12 and 50 hours, respectively. The end-to-end prediction time for a pre-processed MPRAGE volume (~ 256x256x210) was approximately 12 seconds, with 4 seconds for synthesizing WMn-MPRAGE images and 8 seconds for segmenting the thalamus and its nuclei.

Table 2 compares the changes in segmentation accuracy of the NCS CNN based on the choice of loss function on test cohort 1 (n=12 healthy subjects). Although the segmentation CNN trained using WCCE loss function had a higher Dice score for larger nuclei such as Pul, MD, and VLp compared to multi-label Dice loss function, it had significantly higher VD in smaller nuclei such as LGN and MGN ($P<.008$). For the rest of the experimental analyses, NCS and SCS CNNs were trained using multi-label Dice loss function.

Synthesis CNN Performance

Synthesized WMn-MPRAGE images generated from the contrast synthesis CNN are compared to input MPRAGE and target WMn-MPRAGE images in Figure 3. The contrast synthesis CNN is able to mimic native WMn-MPRAGE contrast, especially between thalamus and white matter and inside the thalamus. A quantitative evaluation of synthesis CNN (Table 3) shows that synthesized WMn-MPRAGE images have a high structural similarity (93.7% and 92.1%) and PSNR (23.5 and 21.3) on both test cohorts. Joint histograms between co-registered MPRAGE, WMn-MPRAGE, and synthesized WMn-MPRAGE images for the thalamus region from two test images (Supplementary Figure 1) show high concordance between histograms of WMn-MPRAGE images and synthesized WMn-MPRAGE images.

Segmentation CNN Performance

A qualitative comparison of the thalamic nuclei segmentation predictions from NCS and SCS CNNs for a test subject are shown in Figure 4. Note the improved intra-thalamic contrast as well as thalamus-WM boundary in SCS (arrows) compared to NCS. Both the CNN models are able to accurately identify thalamic nuclei relative to the ground truth for the larger nuclei. However, the boundaries of the smaller nuclei such as VA, Hb, and MTT (Figure 4) and AV, VL_a, CM, LGN, MGN (Figure 5) from SCS CNN agree better with the ground truth annotations.

On the AUD cohort (Table 4), the segmentation CNNs yielded Dice scores of at least 0.84 for the larger nuclei (Pul, MD, VLp) and at least 0.7 for the smaller nuclei. We also observed that synthesizing WMn-MPRAGE representation prior to segmentation significantly reduced overall VD of the whole thalamus ($P=.008$). For nuclei such as MGN ($P=.003$) and CM ($P=.01$), the predictions from SCS CNN yielded a significantly higher Dice score (MGN, $P=.003$ and CM, $P=.01$). Although the Dice scores were relatively comparable, SCS CNN made smaller errors in volume estimates compared to NCS in most of the nuclei, with significant lower error in volume estimates for VPI ($P=.01$) and VA ($P=.001$). Similar trends were also observed in test cohort 1 (Supplementary Table 2) where segmentation using synthesized WMn-MPRAGE representation instead of MPRAGE images yielded

significantly improved Dice scores (AV and CM, $P < .005$) and VD (CM, $P < .005$; MD, $P < .01$) for some nuclei and comparable performance in others.

Effect of Diagnosis on Thalamic Nuclei Volumes in AUD Cohort

The least-squares mean (and standard error) from the ANCOVA analysis on the AUD cohort are presented in Table 5. There was a significant effect of diagnosis on lateral nuclei group volumes after controlling for age and ICV. This reduction in lateral group of nuclei volumes in alcoholics was observed in volumes obtained from both SCS ($F [1, 39] = 6.96, p = 0.01$) and NCS CNNs ($F [1, 39] = 6.92, p = 0.01$). No significant effect of diagnosis was observed on other nuclei groups (anterior, medial, or posterior). Among the constituent nuclei of the lateral group (VLp, VPI, VA, VLa), both SCS ($F [1, 39] = 6.68, p = 0.01$) and NCS CNNs ($F [1, 39] = 5.96, p = 0.04$) segmentations showed a significant atrophy in VLp nucleus volumes in alcoholism. Note that these results are consistent with the previously reported study by Zahr et al (2020) where a significant atrophy was observed in the VLp nucleus in alcoholics but using WMn-MPRAGE images segmented using THOMAS. In addition to VLp, we noticed that the NCS CNN also showed a significant reduction in VPI ($F [1, 39] = 7.92, p = 0.008$) nucleus for subjects with AUD. However, neither Zahr et al (2020) nor the SCS CNN framework showed any significant atrophy in the VPI nucleus.

Bland-Altman plots to assess agreement in CNN-predicted volumes and the ground truth for the whole thalamus, VLp, and the VPI nuclei for the AUD cohort are shown in Figure 6. A stronger correlation between ground truth and volume estimates from SCS CNN is observed compared to NCS CNN. The limits of agreement were also tighter when using synthesized WMn-MPRAGE images for the whole thalamus (Fig. 6d) and VPI nucleus (Fig. 6f) compared to using MPRAGE images (Fig. 6a and Fig. 6c). Bland-Altman plots for NCS CNN show a higher CV for the thalamus (7% vs 4% from SCS CNN) and VPI nucleus (17% vs 10% from SCS CNN).

Figure 7 (and Supplementary Table 3) also compares Dice scores and VD from the NCS CNN trained on WMn-MPRAGE contrast to demonstrate the improvements in segmentation accuracy if true WMn-MPRAGE images were used for training the segmentation CNN. We see that although synthesized WMn-MPRAGE images offer an improvement over MPRAGE contrast, true WMn-MPRAGE contrast still yielded better performance in terms of improved Dice scores and reduced VD for all thalamic nuclei (Figure 7). However, in scenarios where acquisition of WMn-MPRAGE is not possible, such as existing clinical protocols and online repositories, the use of synthesized WMn-MPRAGE can provide improvements over MPRAGE based thalamic nuclei segmentation.

DISCUSSION

In this work, a novel deep learning based thalamic nuclei segmentation framework for MPRAGE images was proposed. A 3D segmentation CNN trained on MPRAGE images (NCS) was developed and the effect of learning a WMn-MPRAGE representation of MPRAGE images in improving the segmentation performance was investigated. The NCS CNN and the segmentation CNN trained on synthesized WMn-MPRAGE images (SCS) were validated on two different cohorts- one containing healthy subjects and one consisting

of a mix of healthy controls and subjects with AUD. Our results show that CNN based segmentation can provide fast and registration-free thalamic nuclei predictions on MPRAGE images. Furthermore, with the training data used in this work, the transformation of MPRAGE images to synthesized WMn-MPRAGE representation was able to improve the accuracy of the segmentation CNN for some nuclei. Figure 8 illustrates an example AUD test subject where the SCS CNN improved the prediction of individual thalamic nuclei in the presence of thalamic lesions. These lesions appear brighter in the synthesized WMn-MPRAGE contrast. Although both the CNN frameworks are trained on healthy individuals, the SCS CNN is able to better delineate thalamic nuclei boundaries, especially of Pul, CM, and VPI nuclei, even in the presence of thalamic lesions.

WMn-MPRAGE Contrast as an Efficient Representation for Thalamic Nuclei Segmentation

Although both the segmentation CNNs showed comparable performance on the thalamus and larger thalamic nuclei on the AUD cohort, the use of SCS CNN was able to significantly improve the Dice scores (MGN and CM) and volume difference (VA and VPI) for some nuclei compared to NCS CNN, and replicate previously observed thalamic nuclear atrophy trends in AUD patients. Both the NCS and the SCS CNNs were able to replicate atrophy in VLp nuclei for subjects with AUD (Zahr et al. 2020) but the NCS CNN showed an additional spurious atrophy in VPI volumes. Note that this is consistent with the significantly higher VD in Table 4 ($13.3\% \pm 12.7\%$ in NCS vs $8.1\% \pm 6.2\%$ in SCS, $P < .008$) and larger standard error in Table 5 for in VPI nucleus. The Bland-Altman plots (Figure 6) reveal that NCS under-estimated volumes from subjects with AUD, artificially inflating the group differences in subjects with thalamic nuclei atrophy due to the underlying pathology. Together, these explain the spurious atrophy in VPI nucleus volumes observed in alcoholics when using volumes computed from the MPRAGE images (i.e. NCS). By transforming the MPRAGE images to a new representation for the segmentation CNN, we were able to demonstrate atrophy only in the VLp nucleus, as previously reported by Zahr et al (2020) using THOMAS and true WMn-MPRAGE data, attesting to the sensitivity, accuracy, and potential utility of representational learning in the segmentation framework used in this work.

Inference Time Computational Efficiency and Existing literature

Manual annotations for the thalamus and thalamic nuclei on structural MR images, although considered gold-standard, are tedious, often requiring hours for accurate delineation. Automated thalamic nuclear segmentation techniques on MRI provide a promising alternative to time-consuming manual annotations. Although recent techniques such as the Bayesian probabilistic atlas-based thalamus segmentation (Iglesias et al. 2018) available in Freesurfer have been proposed for MPRAGE images, the processing time for Freesurfer-based segmentation is on the order of several hours. The shape-based multi-atlas technique proposed by Liu et al (2020) does not take advantage of inter-thalamic nuclear contrast and the segmentation accuracy can be affected in the presence of thalamic nuclear atrophy due to neurological diseases. Moreover, these existing atlas-based segmentation techniques on structural MR images (Iglesias et al. 2018; Su et al. 2019; Liu et al. 2020) rely on success of the nonlinear registration process to accurately transfer thalamic nuclei labels from atlas space to native space. In addition to being computationally burdensome with

longer processing times, the accuracy of the segmentation is often limited by the success of the registration process. The CNN-based segmentation frameworks proposed in this work provide a fully automated, fast, and reliable technique for thalamic parcellations (11 nuclei and MTT) that neither requires registration to a template during inference time nor cropping of the thalamus region to predict individual thalamic nuclei. With prediction times in the order of seconds for a 3D volume, this is a significant reduction compared to multi-atlas schemes like THOMAS (15 min) or probabilistic atlas-based schemes like Freesurfer (several hours).

Recently, a multi-planar 2D CNN-based method was proposed for thalamic nuclei segmentation on WMn-MPRAGE images, where Majdi et al (2020) demonstrated preliminary results of thalamic nuclei segmentation from MPRAGE images using training from WMn-MPRAGE and MPRAGE contrasts. A cascade of 2D CNNs first predicted the thalamus and subsequently the individual thalamic nuclei. The input to the CNNs were cropped images obtained by registering them to a bounding box for thalamus on a WMn-MPRAGE template. This 2D CNN framework was shown to significantly outperform Freesurfer-based segmentation in terms of Dice scores, where the Freesurfer-based segmentation yielded significantly lower Dice scores for small nuclei such as AV, VL_a, CM as well as LGN and MGN (Majdi et al. 2020). As opposed to a cascade of 2D CNNs in Majdi et al (2020), our proposed segmentation framework uses a single 3D CNN to directly predict the thalamus and individual nuclei on MPRAGE images in one pass through the CNN. Furthermore, NCS framework does not require any time-consuming registration-based preprocessing step for cropping.

Limitations and Future Work

A limitation of our work is the lack of manual annotations to train the segmentation CNNs. Manual annotations by experts, although considered gold standard, can be impractical to obtain for a large cohort such as ours. In this work, THOMAS, a multi-atlas framework that uses non-linear registration on WMn-MPRAGE images, is used to generate ground truth labels for training the CNN. The choice of target labels for the thalamus and individual thalamic nuclei, especially one generated from another automated segmentation technique, may affect the segmentation accuracy. However, THOMAS generated nuclei labels have previously been validated against manual annotations (Su et al. 2019). Evaluated on healthy volunteers and patients with Multiple Sclerosis by comparing against manual annotations, THOMAS achieved high Dice scores of at least 0.85 for the larger nuclei and at least 0.7 for smaller structures. Su et al (2019) also compared structural-MRI based THOMAS and Freesurfer-based thalamic nuclei segmentation to manual annotations on a small subset of cases and found that THOMAS outperformed Freesurfer in terms of Dice score for smaller nuclei such as CM, AV, and VA as well as in larger nuclei such as VL_p and Pul. With improved performance against benchmarks, THOMAS-generated labels act as excellent surrogates for time-consuming manual annotations.

Although the segmentation CNNs proposed in this work yield registration-free predictions during inference time, the impact of the success of registration process in the THOMAS label-generation process on CNN predictions would need to be further investigated. Future

studies may have to compare THOMAS and CNN generated labels with manual annotations, if and when available, to understand how errors propagate when training on labels generated by an automated algorithm instead of manual contours. The overall segmentation accuracy of the frameworks can be affected and further improved by the choice of CNN architectures, such as the use of CNN ensembles (Li et al. 2018; Umapathy et al. 2021) for improved generalization performance, or generative adversarial networks (Yi, Walia, and Babyn 2019) for contrast synthesis.

In this study, the segmentation frameworks were trained on a normal cohort and evaluated in subjects with AUD. Improvements in segmentation performance of the CNN was observed when training on a WMn-MPRAGE representation of the MPRAGE images, leading to improved sensitivity to the effects of AUD on thalamic nuclear atrophy on this particular cohort. The clinical applicability of the proposed technique to other neurological diseases that differentially affect thalamic nuclei volumes as well in the presence of lesions or metal electrode artifacts (for e.g. from deep brain stimulation leads) needs further investigation. Online repositories such as ADNI or OASIS with MPRAGE images can be used to explore the relationship between thalamic nuclei volumes and cognition. A fast and reliable MPRAGE-based thalamic parcellation technique such as ours would allow retrospective analysis of these previously acquired neuroimaging data.

CONCLUSION

In this work, the use of a deep learning-based 3D CNN framework was proposed for fast, automated, and registration-free thalamic nuclei segmentation on T1-weighted MPRAGE images. The potential of learning a new representation of MPRAGE images (WMn-MPRAGE contrast) to further improve the performance of segmentation CNN was also investigated. Faster prediction times compared to traditional thalamic nuclei parcellation techniques and the same atrophy in ventral lateral posterior nucleus observed previously with WMn-MPRAGE images with THOMAS segmentation was achieved on subjects with alcoholism using the proposed deep learning frameworks, attesting to their sensitivity, accuracy, as well as clinical utility.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Funding

The authors would like to acknowledge funding from the National Institutes of Health (Grant: R21AA023582) and the Arizona Alzheimer's Disease Consortium as well as support from the Technology and Research Initiative Fund (TRIF) Improving Health Initiative.

REFERENCES

- Aggleton John P, Pralus Agathe, Nelson Andrew J.D., and Hornberger Michael. 2016. "Thalamic Pathology and Memory Loss in Early Alzheimer's Disease: Moving the Focus from the Medial Temporal Lobe to Papez Circuit." *Brain*. 10.1093/brain/aww083.
- Battistella Giovanni, Najdenovska Elena, Maeder Philippe, Ghazaleh Naghme, Daducci Alessandro, Thiran Jean Philippe, Jacquemont Sébastien, et al. 2017. "Robust Thalamic Nuclei Segmentation

- Method Based on Local Diffusion Magnetic Resonance Properties.” *Brain Structure and Function* 222 (5): 2203–16. 10.1007/s00429-016-1336-4. [PubMed: 27888345]
- Behrens TEJ, Johansen-Berg H, Woolrich MW, Smith SM, Wheeler-Kingshott CAM, Boulby PA, Barker GJ, et al. 2003. “Non-Invasive Mapping of Connections between Human Thalamus and Cortex Using Diffusion Imaging.” *Nature Neuroscience* 6 (7). 10.1038/nn1075.
- Bernstein Adam S, Rapcsak Steven Z, Hornberger Michael, Saranathan Manojkumar, and the Alzheimer’s Disease Neuroimaging Initiative. 2021. “Structural Changes in Thalamic Nuclei Across Prodromal and Clinical Alzheimer’s Disease.” *Journal of Alzheimer’s Disease* 82: 361–71. 10.3233/JAD-201583.
- Braak H, and Braak E. 1991. “Alzheimer’s Disease Affects Limbic Nuclei of the Thalamus.” *Acta Neuropathologica* 81 (3): 261–68. 10.1007/BF00305867. [PubMed: 1711755]
- Buckner Randy L., Head Denise, Parker Jamie, Fotenos Anthony F., Marcus Daniel, Morris John C., and Snyder Abraham Z.. 2004. “A Unified Approach for Morphometric and Functional Data Analysis in Young, Old, and Demented Adults Using Automated Atlas-Based Head Size Normalization: Reliability and Validation against Manual Measurement of Total Intracranial Volume.” *NeuroImage* 23 (2). 10.1016/j.neuroimage.2004.06.018.
- Byne W, Buchsbaum MS, Kemether E, Hazlett EA, Shinwari A, Mitropoulou V, and Siever LJ. 2001. “Magnetic Resonance Imaging of the Thalamic Mediodorsal Nucleus and Pulvinar in Schizophrenia and Schizotypal Personality Disorder.” *Archives of General Psychiatry* 58 (2): 133–40. 10.1001/archpsyc.58.2.133. [PubMed: 11177115]
- Conte Gian Marco, Weston Alexander D, Vogelsang David C, Philbrick Kenneth A, Cai Jason C, Barbera Maurizio, Sanvito Francesco, et al. 2021. “Generative Adversarial Networks to Synthesize Missing T1 and FLAIR MRI Sequences for Use in a Multisequence Brain Tumor Segmentation Model.” *Radiology*, March, 203786. 10.1148/radiol.2021203786.
- Dar Salman U.H., Yurt Mahmut, Karacan Levent, Erdem Aykut, Erdem Erkut, and Cukur Tolga. 2019. “Image Synthesis in Multi-Contrast MRI with Conditional Generative Adversarial Networks.” *IEEE Transactions on Medical Imaging* 38 (10): 2375–88. 10.1109/TMI.2019.2901750. [PubMed: 30835216]
- Fama Rosemary, and Sullivan Edith V. 2015. “Thalamic Structures and Associated Cognitive Functions: Relations with Age and Aging.” *Neuroscience & Biobehavioral Reviews* 54: 29–37. <https://doi.org/10.1016/j.neubiorev.2015.03.008>. [PubMed: 25862940]
- Finck Tom, Li Hongwei, Grundl Lioba, Eichinger Paul, Bussas Matthias, Mühlau Mark, Menze Bjoern, and Wiestler Benedikt. 2020. “Deep-Learning Generated Synthetic Double Inversion Recovery Images Improve Multiple Sclerosis Lesion Detection.” *Investigative Radiology* 55 (5). https://journals.lww.com/investigativeradiology/Fulltext/2020/05000/Deep_Learning_Generated_Synthetic_Double_Inversion.9.aspx.
- Hagiwara A, Otsuka Y, Hori M, Tachibana Y, Yokoyama K, Fujita S, Andica C, et al. 2019. “Improving the Quality of Synthetic FLAIR Images with Deep Learning Using a Conditional Generative Adversarial Network for Pixel-by-Pixel Image Translation.” *American Journal of Neuroradiology* 40 (2): 224 LP–230. 10.3174/ajnr.A5927. [PubMed: 30630834]
- Henderson JM, Carpenter K, Cartwright H, and Halliday GM. 2000. “Loss of Thalamic Intralaminar Nuclei in Progressive Supranuclear Palsy and Parkinson’s Disease: Clinical and Therapeutic Implications.” *Brain*. Vol. 123.
- Goodfellow Ian and Bengio Yoshua and Courville Aaron. 2016. “Deep Learning.” In *Deep Learning*, 524–54. MIT Press. <http://www.deeplearningbook.org>.
- Iglehart Charles, Monti Martin, Cain Joshua, Tourdias Thomas, and Saranathan Manojkumar. 2020. “A Systematic Comparison of Structural-, Structural Connectivity-, and Functional Connectivity-Based Thalamus Parcellation Techniques.” *Brain Structure and Function* 225 (5): 1631–42. 10.1007/s00429-020-02085-8. [PubMed: 32440784]
- Iglesias Juan Eugenio, Insausti Ricardo, Lerma-Usabiaga Garikoitz, Bocchetta Martina, Van Leemput Koen, Greve Douglas N., van der Kouwe Andre, Fischl Bruce, Caballero-Gaudes César, and Paz-Alonso Pedro M.. 2018. “A Probabilistic Atlas of the Human Thalamic Nuclei Combining Ex Vivo MRI and Histology.” *NeuroImage* 183 (December): 314–26. 10.1016/j.neuroimage.2018.08.012. [PubMed: 30121337]

- Li Hongwei, Jiang Gongfa, Zhang Jianguo, Wang Ruixuan, Wang Zhaolei, Zheng Wei Shi, and Menze Bjoern. 2018. "Fully Convolutional Network Ensembles for White Matter Hyperintensities Segmentation in MR Images." *NeuroImage* 183: 650–65. 10.1016/j.neuroimage.2018.07.005. [PubMed: 30125711]
- Litjens Geert, Kooi Thijs, Bejnordi Babak Ehteshami, Setio Arnaud Arindra Adiyoso, Ciompi Francesco, Ghafoorian Mohsen, van der Laak Jeroen A.W.M., van Ginneken Bram, and Sánchez Clara I. 2017. "A Survey on Deep Learning in Medical Image Analysis." *Medical Image Analysis* 42 (December 2012): 60–88. 10.1016/j.media.2017.07.005. [PubMed: 28778026]
- Liu Yuan, D'Haese Pierre-François, Newton Allen T, and Dawant Benoit M. 2020. "Generation of Human Thalamus Atlases from 7 T Data and Application to Intrathalamic Nuclei Segmentation in Clinical 3 T T1-Weighted Images." *Magnetic Resonance Imaging* 65: 114–28. <https://doi.org/10.1016/j.mri.2019.09.004>. [PubMed: 31629074]
- Majdi Mohammad S., Keerthivasan Mahesh B., Rutt Brian K., Zahr Natalie M., Rodriguez Jeffrey J., and Saranathan Manojkumar. 2020. "Automated Thalamic Nuclei Segmentation Using Multi-Planar Cascaded Convolutional Neural Networks." *Magnetic Resonance Imaging* 73. 10.1016/j.mri.2020.08.005.
- Olut Sahin, Sahin Yusuf H, Demir Ugur, and Unal Gozde. 2018. "Generative Adversarial Training for MRA Image Synthesis Using Multi-Contrast MRI BT - PRedictive Intelligence in MEDicine." In , edited by Rekek Islem, Unal Gozde, Adeli Ehsan, and Park Sang Hyun, 147–54. Cham: Springer International Publishing.
- Planche Vincent, Su Jason H., Mournet Sandy, Saranathan Manojkumar, Dousset Vincent, Han May, Rutt Brian K., and Tourdias Thomas. 2019. "White-Matter-Nullled MPRAGE at 7T Reveals Thalamic Lesions and Atrophy of Specific Thalamic Nuclei in Multiple Sclerosis." *Multiple Sclerosis Journal*. 10.1177/1352458519828297.
- Ronneberger Olaf, Fischer Philipp, and Brox Thomas. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation," May. <http://arxiv.org/abs/1505.04597>.
- Rosa Francesco La, Yu Thomas, Barquero Germán, Thiran Jean-Philippe, Granziera Cristina, and Cuadra Meritxell Bach. 2021. "MPRAGE to MP2RAGE UNI Translation via Generative Adversarial Network Improves the Automatic Tissue and Lesion Segmentation in Multiple Sclerosis Patients." *Computers in Biology and Medicine* 132: 104297. <https://doi.org/10.1016/j.combiomed.2021.104297>. [PubMed: 33711559]
- Saranathan Manojkumar, Tourdias Thomas, Bayram Ersin, Ghanouni Pejman, and Rutt Brian K.. 2015. "Optimization of White-Matter-Nullled Magnetization Prepared Rapid Gradient Echo (MP-RAGE) Imaging." *Magnetic Resonance in Medicine* 73 (5): 1786–94. 10.1002/mrm.25298. [PubMed: 24889754]
- Simonyan Karen, and Zisserman Andrew. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, 1–14.
- Smith Stephen M. 2002. "Fast Robust Automated Brain Extraction." *Human Brain Mapping* 17 (3): 143–55. 10.1002/hbm.10062. [PubMed: 12391568]
- Su Jason H., Thomas Francis T., Kasoff Willard S., Tourdias Thomas, Choi Eun Young, Rutt Brian K., and Saranathan Manojkumar. 2019. "Thalamus Optimized Multi Atlas Segmentation (THOMAS): Fast, Fully Automated Segmentation of Thalamic Nuclei from Structural MRI." *NeuroImage* 194 (October 2018): 272–82. 10.1016/j.neuroimage.2019.03.021. [PubMed: 30894331]
- Tourdias Thomas, Saranathan Manojkumar, Levesque Ives R., Su Jason, and Rutt Brian K.. 2014. "Visualization of Intra-Thalamic Nuclei with Optimized White-Matter-Nullled MPRAGE at 7T." *NeuroImage* 84: 534–45. 10.1016/j.neuroimage.2013.08.069. [PubMed: 24018302]
- Tustison Nicholas J., Avants Brian B., Cook Philip A., Zheng Yuanjie, Egan Alexander, Yushkevich Paul A., and Gee James C.. 2010. "N4ITK: Improved N3 Bias Correction." *IEEE Transactions on Medical Imaging* 29 (6): 1310–20. 10.1109/TMI.2010.2046908. [PubMed: 20378467]
- Umapathy Lavanya, Perez-Carrillo Gloria Guzman, Keerthivasan Mahesh Bharath, Rosado-Toro José A, Altbach Maria I, Winegar Blair, Weinkauff Craig, Bilgin Ali, and for the Alzheimer's Disease Neuroimaging Initiative. 2021. "A Stacked Generalization of 3D Orthogonal Deep Learning Convolutional Neural Networks for Improved Detection of White Matter Hyperintensities in 3D FLAIR Images." *American Journal of Neuroradiology* In pres. <https://doi.org/10.3174/ajnr.A6970>.

- Wang Tonghe, Lei Yang, Fu Yabo, Wynne Jacob F, Curran Walter J, Liu Tian, and Yang Xiaofeng. 2021. "A Review on Medical Imaging Synthesis Using Deep Learning and Its Clinical Applications." *Journal of Applied Clinical Medical Physics* 22 (1): 11–36. <https://doi.org/10.1002/acm2.13121>.
- Wei Wen, Poirion Emilie, Bodini Benedetta, Durrleman Stanley, Colliot Olivier, Stankoff Bruno, and Ayache Nicholas. 2019. "Fluid-Attenuated Inversion Recovery MRI Synthesis from Multisequence MRI Using Three-Dimensional Fully Convolutional Networks for Multiple Sclerosis." *Journal of Medical Imaging* 6 (1): 1–9. 10.1117/1.JMI.6.1.014005.
- Wiegell Mette R., Tuch David S., Larsson Henrik B.W., and Wedeen Van J.. 2003. "Automatic Segmentation of Thalamic Nuclei from Diffusion Tensor Magnetic Resonance Imaging." *NeuroImage* 19 (2): 391–401. 10.1016/S1053-8119(03)00044-2. [PubMed: 12814588]
- Yang Qianye, Li Nannan, Zhao Zixu, Fan Xingyu, Chang Eric I-Chao, and Xu Yan. 2020. "MRI Cross-Modality Image-to-Image Translation." *Scientific Reports* 10 (1): 3753. 10.1038/s41598-020-60520-6. [PubMed: 32111966]
- Yi Xin, Walia Ekta, and Babyn Paul. 2019. "Generative Adversarial Network in Medical Imaging: A Review." *Medical Image Analysis* 58: 101552. <https://doi.org/10.1016/j.media.2019.101552>. [PubMed: 31521965]
- Zahr Natalie M., Sullivan Edith V., Pohl Kilian M., Pfefferbaum Adolf, and Saranathan Manojkumar. 2020. "Sensitivity of Ventrolateral Posterior Thalamic Nucleus to Back Pain in Alcoholism and CD4 Nadir in HIV." *Human Brain Mapping* 41 (5): 1351–61. 10.1002/hbm.24880. [PubMed: 31785046]

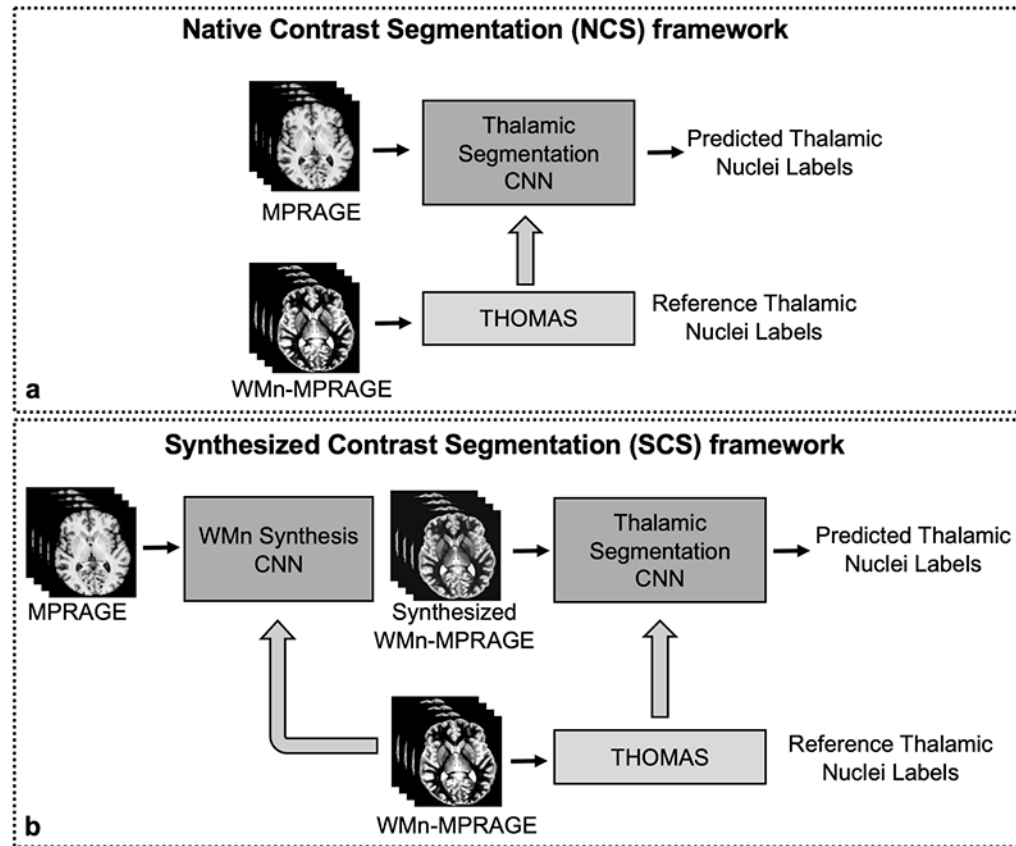


Fig. 1: Overview of the proposed multi-scale thalamic nuclei segmentation frameworks. (a) The thalamus and its individual nuclei are predicted directly on MPRAGE images using the Native Contrast Segmentation (NCS) framework. (b) A contrast synthesis CNN is introduced prior to the segmentation CNN for synthesized contrast segmentation (SCS) framework. This synthesis CNN learns a new representation (WMn-MPRAGE) from MPRAGE contrast. The ground truth labels are generated on the WMn-MPRAGE images using a multi-atlas label fusion algorithm (THOMAS).

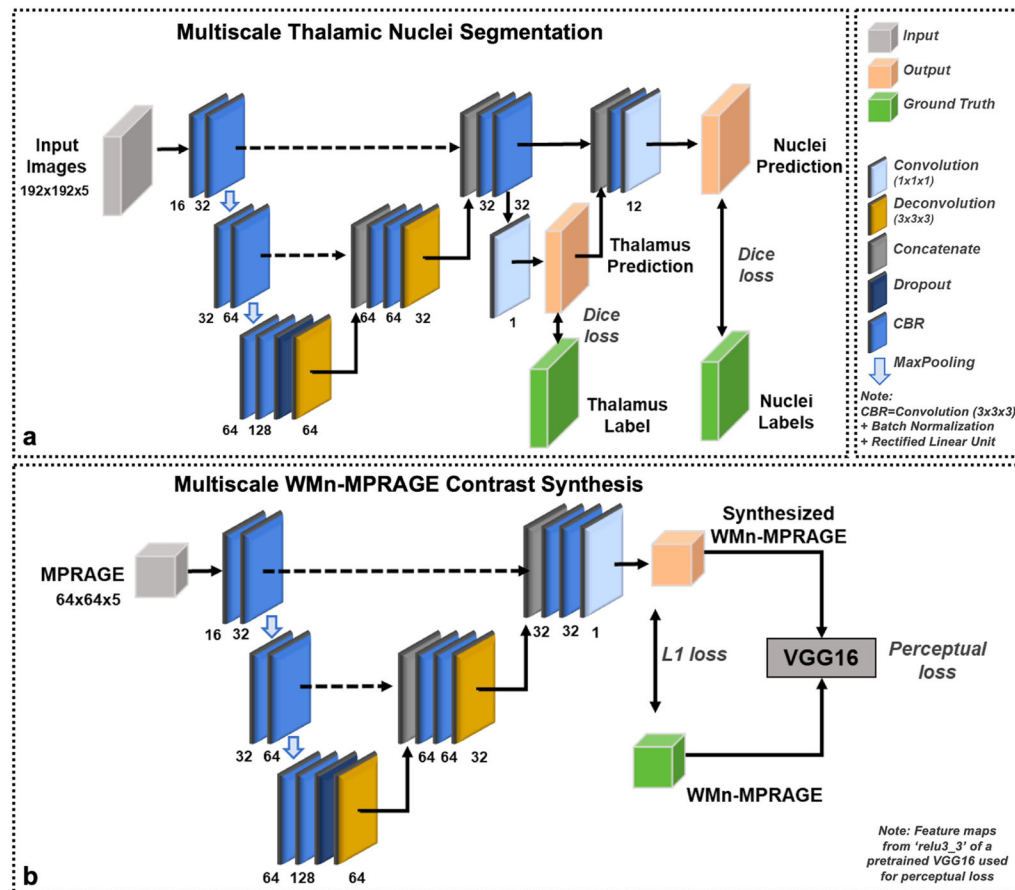


Fig. 2: Convolutional Neural Network (CNN) architectures for synthesis and segmentation. (a) A 3D multi-scale thalamic nuclei segmentation CNN and (b) a 3D multi-scale white-matter-nulled contrast synthesis CNN are shown. The number of feature maps generated from each layer are denoted next to the layer. Note that both architectures have a similar encoder-decoder like framework with skip connections except for the choice of loss functions.

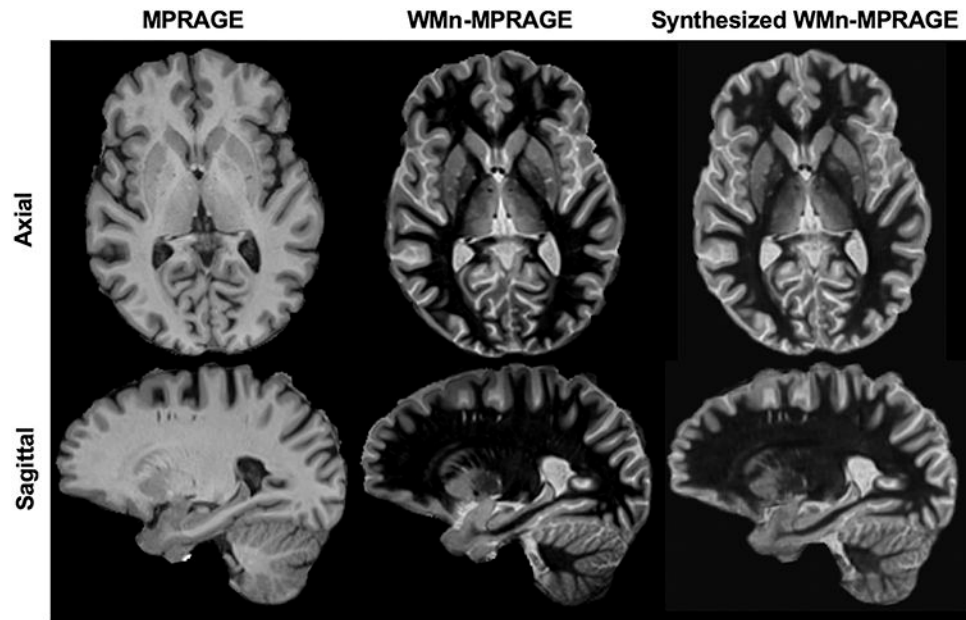


Fig. 3: White-matter-nulled MPRAGE (WMn-MPRAGE) synthesis. Representative axial and sagittal cross-sections of a synthesized image from a test subject are shown. The target WMn-MPRAGE images as well as input MPRAGE images are also shown for reference. The synthesized WMn-MPRAGE images are able to mimic the contrast between thalamus and the surrounding white matter similar to the WMn-MPRAGE images.

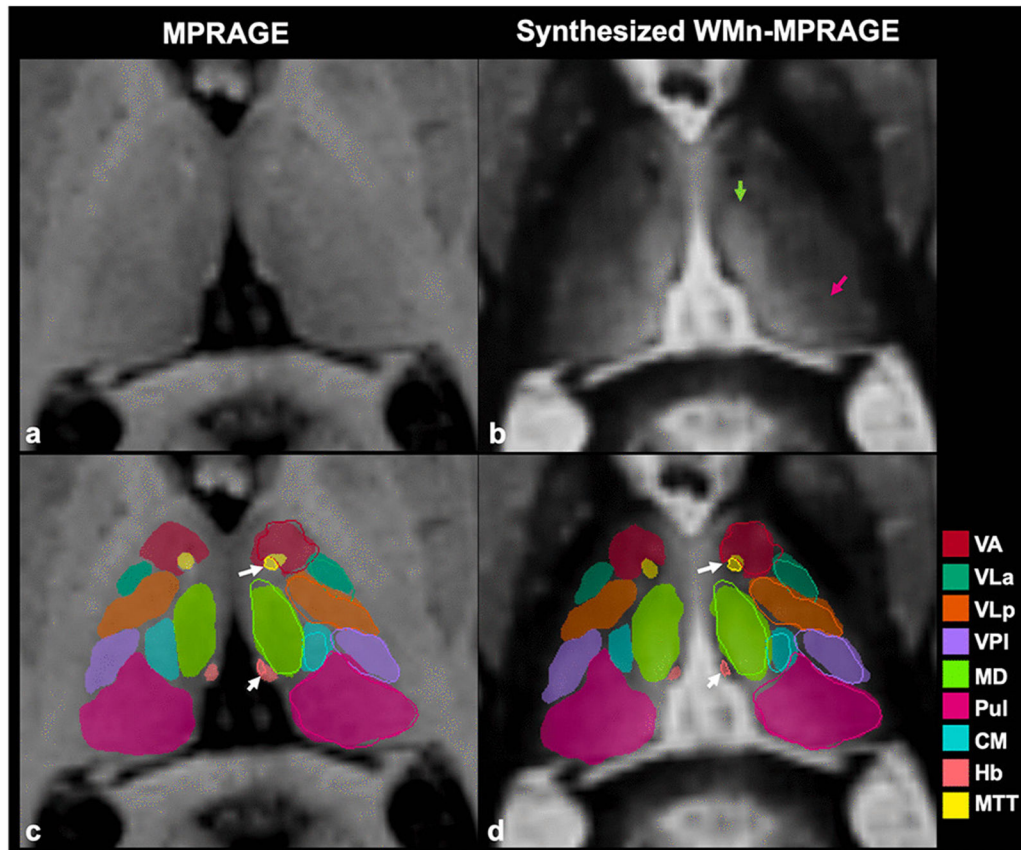


Fig. 4: Thalamic nuclei prediction on MPRAGE and synthesized WMn-MPRAGE images using NCS and the SCS CNN frameworks, respectively. Top row (a-b): Representative axial cross-sections from a test subject for the different image contrast are shown. It is to be noted that synthesized WMn-MPRAGE contrast shows improved inter-thalamic contrast. The arrows in (b) highlight nuclei boundaries on synthesized WMn-MPRAGE image for MD and Pul nuclei. Bottom row: Individual thalamic nuclei labels predictions from the (c) MPRAGE contrast and (d) synthesized WMn-MPRAGE contrast are overlaid on the images with reference THOMAS labels outlined on the left thalamus. The segmentation CNN using synthesized contrast is able to accurately delineate even smaller nuclei such as ventral anterior (VA), habenula (Hb), and mammillothalamic tract (MTT).

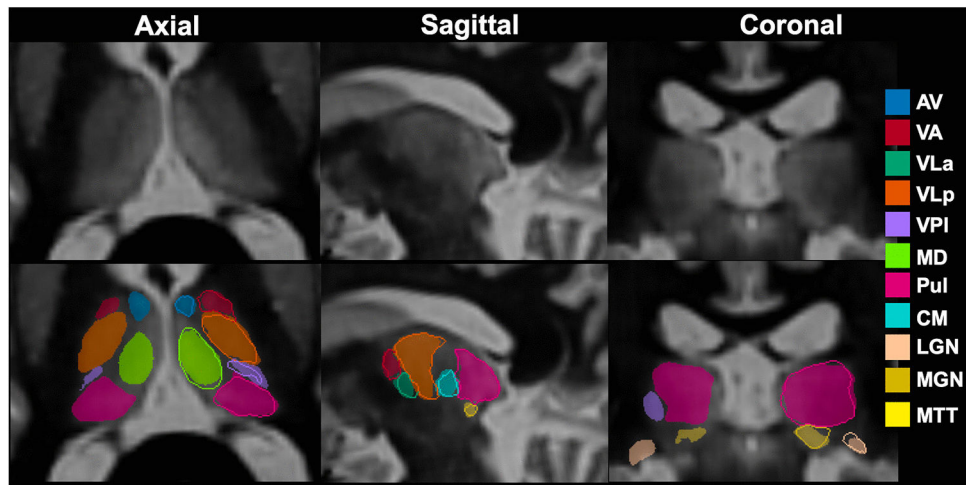


Fig. 5: Thalamic nuclei predictions from the synthesized contrast segmentation CNN. Multiplanar views of the test subject (top row) in Figure 4 with individual nuclei labels from the antral, ventral, posterior, and medial thalamic nuclei groups predicted using synthesized WMn-MPRAGE (bottom row) are shown. The corresponding ground truth labels from THOMAS are overlaid (color outlines) for the left thalamus. The coronal view shows the LGN and MGN nuclei. The predicted labels for the smaller nuclei such as AV, VA, VLp, CM, LGN, and MGN agree well with the THOMAS outlines.

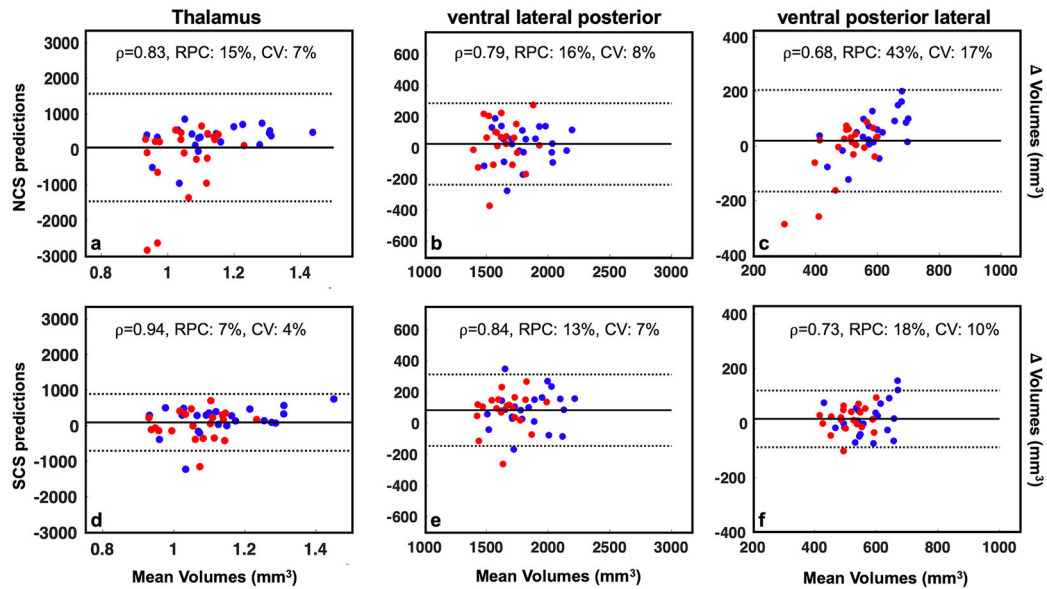


Fig. 6:

Bland-Altman plots to assess agreement between predicted volumes and ground truth THOMAS labels on the Alcohol Use Disorder (AUD) cohort. Bland-Altman plots for agreement in volumes for Thalamus (a), ventral lateral posterior (b), and ventral posterior lateral nucleus (c) for the segmentation CNN trained on MPRAGE images (top row) are shown. The plots are also generated for the volumes predicted by the segmentation CNN trained on synthesized WMn-MPRAGE images (bottom row). The subjects are color-coded with blue for controls and red for subjects with AUD. For each case, Pearson's correlation coefficient (ρ), percent repeatability coefficient (RPC), and coefficient of variation (CV) are also reported. It should be noted that on the AUD cohort, the volumes from synthesized WMn-MPRAGE images have tighter limits of agreement. Additionally, the NCS framework underestimates the volumes for thalamus and ventral posterior lateral nucleus for subjects with AUD (red).

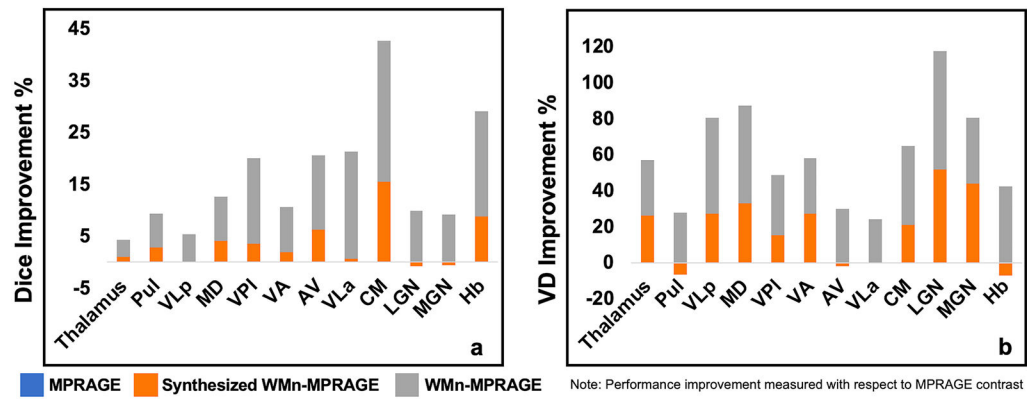


Fig. 7: Comparison of thalamic nuclei segmentation CNN performance with different image contrasts on the test cohorts (test cohort 1 and alcohol use disorder cohort). The bar plots compare the improvements in Dice score (a) and volume difference (VD) (b) for the different nuclei when using synthesized WMn-MPRAGE and WMn-MPRAGE over conventional MPRAGE contrast. The nuclei are arranged in decreasing order of their average volumes. Transforming the MPRAGE contrast to synthesized WMn-MPRAGE representation yields improvements in segmentation performance. The usage of true WMn-MPRAGE contrast outperforms MPRAGE based segmentation.

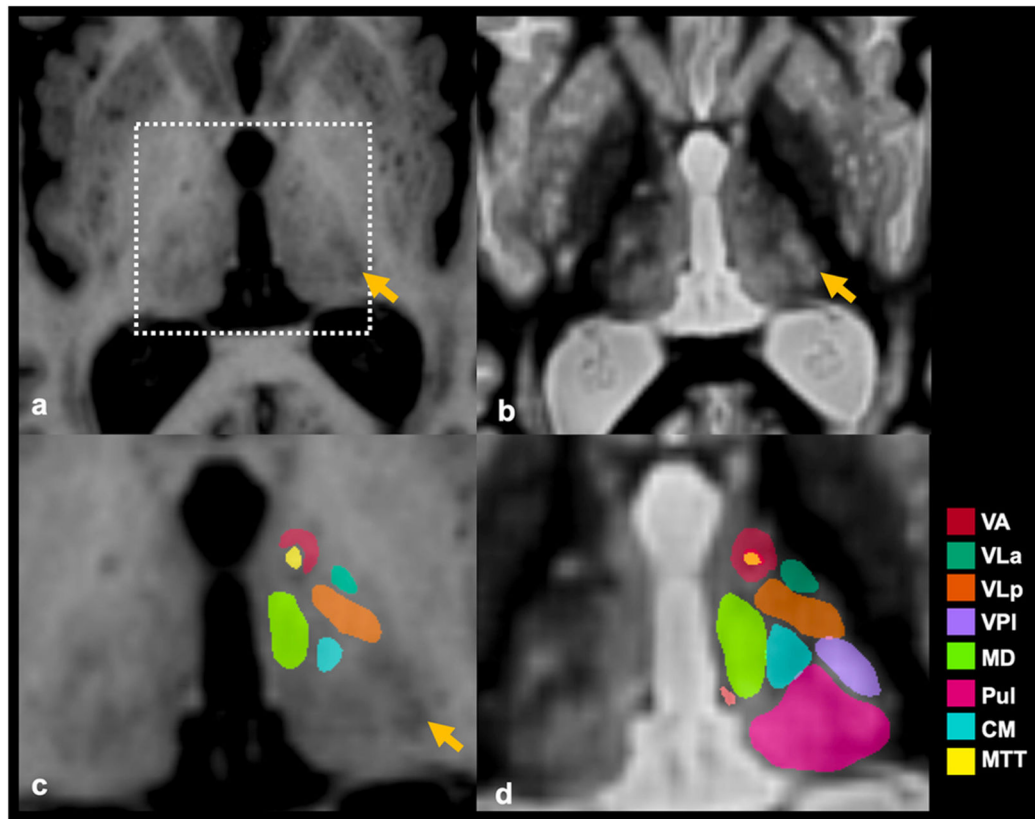


Fig. 8: Thalamic nuclei prediction on a subject with lesions in the thalamus. Axial cross-sections of the subject, with alcohol use disorder, are shown in the top row (a-b). The thalamic lesions (yellow arrows) appear bright in the WM-nulled representation of MPRAGE image. The predictions from the corresponding NCS CNN (c) and SCS CNN (d) are shown. On this particular case, the segmentation using MPRAGE contrast (NCS) fails to correctly delineate nuclei surrounding the lesions, especially centromedian (CM), pulvinar (Pul), and ventral posterior lateral (VPI) nuclei.

Table 1:

Index of thalamic nuclei groups and abbreviations

Groups	Thalamic Nuclei
Anterior	Anterior ventral nucleus (AV)
	Ventral lateral posterior nucleus (VLp)
Lateral	Ventral lateral anterior nucleus (VLa)
	Ventral anterior nucleus (VA)
	Ventral posterior lateral nucleus (VPL)
	Pulvinar nucleus (Pul)
Posterior	Medial geniculate nucleus (MGN)
	Lateral geniculate nucleus (LGN)
	Mediodorsal nucleus (MD)
Medial	Centromedian nucleus (CM)
	Habenular nucleus (Hb)
Others	Mammillothalamic tract (MTT)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Quantitative comparison (Mean \pm SD) of segmentation CNN performance with different loss functions (test cohort 1)

		Dice Scores		Volume Difference (%)	
		WCCE ^a	Multi-label Dice	WCCE	Multi-label Dice
Lateral	Thalamus	0.91 \pm 0.03	0.91 \pm 0.03	7.7 \pm 7.3	6.1 \pm 7.1
	VLp	0.84 \pm 0.03 *	0.82 \pm 0.04	6.4 \pm 5.2	7.3 \pm 7.4
	VPI	0.74 \pm 0.06	0.66 \pm 0.12	16.9 \pm 12.4	17.2 \pm 18.4
	VA	0.723 \pm 0.08	0.72 \pm 0.10	20.3 \pm 10.0	14.0 \pm 13.8
	VLa	0.638 \pm 0.11	0.62 \pm 0.11	45.2 \pm 39.5	23.8 \pm 19.9
	Pul	0.89 \pm 0.02 *	0.84 \pm 0.05	4.7 \pm 3.9 *	10.9 \pm 5.4
Posterior	MGN	0.72 \pm 0.08	0.77 \pm 0.07 *	27.0 \pm 13.3	9.3 \pm 8.0 *
	LGN	0.63 \pm 0.20	0.74 \pm 0.14 *	29.4 \pm 20.4	13.2 \pm 10.6 *
	MD	0.86 \pm 0.07 *	0.80 \pm 0.12	12.5 \pm 13.6	9.0 \pm 10.2
Medial	CM	0.67 \pm 0.06	0.57 \pm 0.17	52.1 \pm 29.2	37.6 \pm 26.3
	Hb	0.26 \pm 0.22	0.56 \pm 0.28 *	48.6 \pm 33.3	25.4 \pm 32.8
	AV	0.70 \pm 0.17	0.66 \pm 0.14	40.2 \pm 24.3	24.0 \pm 21.1

^aWCCE = Weighted Categorical Cross Entropy

*P<0.008, two-sided paired t-test

Table 3:Evaluation (mean \pm SD) of synthesis CNN performance

	Test Cohort 1	AUD Cohort
Root mean squared error (RMSE)	0.067 \pm 0.007	0.086 \pm 0.009
Peak signal to noise (PSNR)	23.5 \pm 0.968	21.3 \pm 0.829
Structural similarity (SSIM)	0.937 \pm 0.010	0.921 \pm 0.017

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4:Quantitative comparison of CNN frameworks (Mean \pm SD) on Alcohol Use Disorder Cohort (n=45)

	Dice Score		Volume Difference (%)	
	NCS ^a	SCS ^b	NCS	SCS
Thalamus	0.92 \pm 0.03	0.93 \pm 0.01	4.9 \pm 5.0	2.9 \pm 2.3 *
VLp	0.84 \pm 0.04	0.84 \pm 0.04	6.4 \pm 5.0	7.1 \pm 4.7
VPI	0.73 \pm 0.12	0.73 \pm 0.1	13.3 \pm 12.7	8.1 \pm 6.2 *
VA	0.77 \pm 0.04	0.78 \pm 0.05	10.8 \pm 8.6	7.3 \pm 6.5 **
VLa	0.65 \pm 0.12	0.67 \pm 0.1	20.1 \pm 16.0	16.8 \pm 12.1
Pul	0.86 \pm 0.09	0.88 \pm 0.03	8.1 \pm 12.2	5.4 \pm 4.6
MGN	0.76 \pm 0.08	0.78 \pm 0.07 **	11.6 \pm 9.7	9.6 \pm 7.1
LGN	0.76 \pm 0.09	0.77 \pm 0.08	15.1 \pm 11.6	13.5 \pm 13.4
MD	0.86 \pm 0.05	0.87 \pm 0.03	9.0 \pm 10.2	6.7 \pm 4.6
CM	0.72 \pm 0.12	0.76 \pm 0.07 *	14.6 \pm 14.5	12.3 \pm 11.2
Hb	0.68 \pm 0.11	0.69 \pm 0.08	17.2 \pm 12.8	14.4 \pm 7.9
AV	0.75 \pm 0.11	0.74 \pm 0.09	26.9 \pm 51.7	24.9 \pm 59.8

* P=.01, two-sided paired t-test

** P<.005, two-sided paired t-test

^aNCS: Native Contrast Segmentation CNN^bSCS: Synthesized Contrast Segmentation CNN

Table 5:Thalamic atrophy in alcoholism: ANCOVA^a analysis on the Alcohol Use Disorder cohort

Lateral Group			
	Controls	AUD	P-value
NCS	3133.1 ± 69.7	2865.2 ± 71.5	0.01*
SCS	3153.3 ± 49.0	2964 ± 50.2	0.01*
Ventral lateral posterior			
	Controls	AUD	P-value
NCS	1787.6 ± 34.9	1678.5 ± 35.8	0.04*
SCS	1850.2 ± 31	1733.2 ± 31.7	0.01*
Ventral posterior lateral			
	Controls	AUD	P-value
NCS	604.1 ± 23.4	507.8 ± 24	0.008*
SCS	567.8 ± 12.8	537.1 ± 13.1	0.1

^aThe values displayed are the least squares mean ± standard error from the ANCOVA analysis (n=45).