



**HAL**  
open science

## Robust thalamic nuclei segmentation from T1-weighted MRI using polynomial intensity transformation

Julie Vidal, Lola Danet, Patrice Péran, Jérémie Pariente, Meritxell Bach Cuadra, Natalie Zahr, Emmanuel Barbeau, Manojkumar Saranathan

► **To cite this version:**

Julie Vidal, Lola Danet, Patrice Péran, Jérémie Pariente, Meritxell Bach Cuadra, et al.. Robust thalamic nuclei segmentation from T1-weighted MRI using polynomial intensity transformation. Brain Structure and Function, In press, 10.1007/s00429-024-02777-5 . hal-04584276

**HAL Id: hal-04584276**

**<https://ut3-toulouseinp.hal.science/hal-04584276v1>**

Submitted on 23 May 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Robust thalamic nuclei segmentation from T1-weighted MRI using polynomial intensity transformation

Julie P. Vidal<sup>1,2</sup>, Lola Danet<sup>2,3</sup>, Patrice Péran<sup>2</sup>, Jérémie Pariente<sup>2,3</sup>, Meritxell Bach Cuadra<sup>4</sup>, Natalie M. Zahr<sup>5,6</sup>, Emmanuel J. Barbeau<sup>1</sup>, and Manojkumar Saranathan<sup>\*7</sup>

<sup>1</sup>CNRS, CerCo (Brain and Cognition Research Center) - Paul Sabatier University, Toulouse, France, <sup>2</sup>INSERM, ToNiC (Toulouse NeuroImaging Center) – Paul Sabatier University, Toulouse, France, <sup>3</sup>Purpan Hospital, Toulouse University Hospital Center, Neurology Department, Toulouse, France, <sup>4</sup>CIBM (BioMedical Imaging Center) – Radiology Department, Lausanne University and University Hospital, Lausanne, Switzerland, <sup>5</sup>Center for Health Sciences, SRI International, Menlo Park, California, USA, <sup>6</sup>Department of Psychiatry & Behavioral Sciences - Stanford University School of Medicine, Stanford, California, USA, <sup>7</sup>Department of Radiology - University of Massachusetts Chan Medical School, Worcester, Massachusetts, USA

\* Email: [Manojkumar.Saranathan@umassmed.edu](mailto:Manojkumar.Saranathan@umassmed.edu)

Under review in Brain structure and function.

## **Abstract**

Accurate segmentation of thalamic nuclei, crucial for understanding their role in healthy cognition and in pathologies, is challenging to achieve on standard T1-weighted (T1w) magnetic resonance imaging (MRI) due to poor image contrast. White-matter-nulled (WMn) MRI sequences improve intrathalamic contrast but are not part of clinical protocols or extant databases. In this study, we introduce histogram-based polynomial synthesis (HIPS), a fast preprocessing transform step that synthesizes WMn-like image contrast from standard T1w MRI using a polynomial approximation for intensity transformation. HIPS was incorporated into Thalamus Optimized Multi-Atlas Segmentation (THOMAS) pipeline, a method developed and optimized for WMn MRI. HIPS-THOMAS was compared to a convolutional neural network (CNN)-based segmentation method and THOMAS modified for T1w images (T1w-THOMAS). The robustness and accuracy of the three methods were tested across different image contrasts (MPRAGE, SPGR, and MP2RAGE), scanner manufacturers (PHILIPS, GE, and Siemens), and field strengths (3T and 7T). HIPS-transformed images improved intra-thalamic contrast and thalamic boundaries, and HIPS-THOMAS yielded significantly higher mean Dice coefficients and reduced volume errors compared to both the CNN method and T1w-THOMAS. Finally, all three methods were compared using the frequently travelling human phantom MRI dataset for inter- and intra-scanner variability, with HIPS displaying the least inter-scanner variability and performing comparably with T1w-THOMAS for intra-scanner variability. In conclusion, our findings highlight the efficacy and robustness of HIPS in enhancing thalamic nuclei segmentation from standard T1w MRI.

**Keywords:** Thalamus, thalamic nuclei segmentation, THOMAS, structural imaging

## 1. Introduction

The thalamus is a deep brain structure on either side of the third ventricle, comprised of multiple nuclei (Morel et al., 1997). Physiologically, these nuclei are often classified as first order nuclei (e.g. lateral and medial geniculate nuclei), relaying sensory or motor information to the cortex, higher order nuclei (e.g. pulvinar) involved in cognition through cortico-thalamo-cortical circuits (Sherman, 2007) and the thalamic reticular nucleus. Several neurodegenerative, neurological, and neuropsychiatric conditions involve the thalamic nuclei such as alcohol use disorder (Zahr et al., 2020), schizophrenia (Andreasen, 1997), Alzheimer's disease (Braak & Braak., 1991), chronic pain syndrome (Gustin et al., 2011), epilepsy (Fisher et al., 2010) and stroke (Danet et al., 2015). Thus, visualization and characterization of thalamic nuclei are crucial in understanding their function as well as their relationship in the healthy brain or in pathological conditions. This need is even more critical for pathologies where clinical care is based on accurate targeting of specific nuclei, such as deep brain stimulation (DBS) for the treatment of essential tremor (Koller et al., 2001), chronic pain syndrome (Owen et al., 2006) or drug-resistant epilepsy (Fisher et al., 2010). Other therapeutic approaches targeting specific thalamic nuclei include Magnetic Resonance Guided Focused Ultrasound thalamotomy (MRgFUS), gamma knife surgery, radio frequency surgery, and microsurgical resection (Benabid et al., 1997; Cinalli et al., 2018).

Segmentation of thalamic nuclei is challenging due to their poor contrast on standard T1 and T2 weighted (T1w, T2w) MRI, the most commonly used pulse sequences in routine neuroimaging protocols. To partly address this issue, atlases such as the Schaltenbrand-Warren and the Morel atlas (Morel et al., 1997; Schaltenbrand & Warren, 1977) have been developed. These atlases have been used for manual delineations of nuclei on MRI data as well as in neurosurgical targeting using stereotactic coordinates (Sanborn et al., 2009). Manual delineation of thalamic nuclei requires special expertise, is time consuming, and is therefore not ideal for analysis of large datasets. In addition, the small size of several nuclei makes them difficult to delineate. A three-dimensional reconstruction of the Morel Atlas in MNI space has been developed to help automate the segmentation of thalamic

nuclei (Jakab et al., 2012). However, such atlas-based approaches ignore important inter-individual and inter-thalamic variability of the size, shape, and location of thalamic nuclei, leading to compromised accuracy.

To address the issue of automated thalamic nuclei parcellation at a subject-level and exploit image contrast, several approaches have been explored. One of the earliest methods, proposed by Behrens et al. (2003), used probabilistic tractography from Diffusion Tensor Imaging (DTI) data, mapping structural connectivity between different thalamic regions and specific cortical regions based on white matter anisotropy (e.g. mediodorsal nucleus to prefrontal cortex, lateral geniculate nucleus to visual cortex and so on), resulting in 7 thalamic regions corresponding to the 7 seed regions. Other diffusion MRI based methods have used local information from the diffusion tensor at a voxel level to parcellate the thalamus. Mang et al. (2012) used k-means clustering of the dominant diffusion orientation while Battistella et al. (2017) used k-means clustering of the spherical harmonic coefficients of the orientation distribution function to parcellate the thalamus into 6 regions. Since the thalamus is mainly composed of isotropic grey matter, the direction information computed from the diffusion tensor tends to be noisy. Further, the spatial resolution limitations of the underlying echo planar imaging acquisition results in a small number of clusters rather than precise, anatomically defined nuclei. Functional MRI based segmentation approaches have also been proposed using the idea of functional connectivity to cortical ROIs employing seed-based (Zhang et al., 2008), or Independent Components Analysis (ICA) (Hale et al., 2015). These also resulted in 6 thalamic regions corresponding to structural connectivity results of Behrens et al. In contrast, to parcellate the thalamus into 15 clusters, Kumar et al. (2017) proposed an ICA based functional parcellation while Van Oort et al. (2018) used time courses of instantaneous connectivity. Other methods based on susceptibility weighted imaging or quantitative susceptibility mapping (Liu et al., 2020; Deoni et al., 2005) have typically relied upon manual segmentation and have been limited to targeting the VIM nucleus for DBS treatment of movement disorders.

Images from structural (anatomic) MRI methods such as T1w Magnetization Prepared Rapid Gradient Echo (MPRAGE) have high spatial resolution (typically 1mm isotropic), minimal distortion, and are commonly used for cortical segmentation but rarely used for thalamic nuclei segmentation due to poor intra-thalamic nuclear contrast. Iglesias et al. developed a probabilistic atlas that combined manual delineations from in vivo T1w MPRAGE data and ex-vivo histological data (Iglesias et al., 2018). The Bayesian segmentation algorithm that uses this atlas to segment T1w MRI is part of the Freesurfer package. Variants of MPRAGE such as white-matter-nulled (WMn) (Tourdias et al., 2014), and grey-matter-nulled (GMn) (Magnotta et al., 2000) MPRAGE imaging significantly improves intra-thalamic contrast and permit better delineation of thalamic nuclei. A multi-atlas method called Thalamus Optimized Multi-Atlas Segmentation (THOMAS) (Su et al., 2019) has been proposed that uses 20 WMn-MPRAGE prior datasets acquired at 7T and segmented manually using the Morel atlas as a guide combined with a joint-label fusion algorithm for thalamic nuclei parcellation of WMn-MPRAGE data. This method (which we call WMn-THOMAS) divided the thalamus into 11 nuclei per hemisphere and was validated against manual segmentation.

While WMn-THOMAS has been used in several studies examining the role of thalamic nuclei in alcohol use disorder and multiple sclerosis (Zahr et al., 2020; Su et al., 2020), WMn-MPRAGE sequences are neither part of commonly used clinical protocols nor available in existing data repositories like ADNI or OASIS. To segment conventional MPRAGE T1w data, THOMAS was recently modified to use mutual information (MI) instead of cross-correlation (CC) as the nonlinear registration metric (Bernstein et al., 2021; Pfefferbaum et al., 2023) and a majority voting algorithm for label fusion (which we call T1w-THOMAS). While this method achieved good accuracy compared to WMn-MPRAGE for larger nuclei such as the mediodorsal or pulvinar, it was less accurate for segmentation of the smaller centromedian and habenular nuclei. One reason for suboptimal performance of the modified T1w-THOMAS method could be due to poor intrathalamic contrast and thalamic boundaries on standard T1w contrast images. To address this limitation, a novel deep learning-based method (Umapathy et al., 2021) was proposed, that first synthesized WMn-MPRAGE-like images from T1w data and then

segmented the synthetic WMn-like data, using two separately trained convolutional neural networks (CNN). This CNN-based method was shown to improve accuracy compared to direct segmentation of T1w data (using a different CNN trained directly on 3T T1w data). However, the adoption of this method has been limited due to the necessity of training the CNNs on new types of data (e.g. different field strengths or scanner manufacturer). Furthermore, this CNN training process is time consuming and often not possible due to lack of T1w and WMn-MPRAGE data acquired concurrently on the same subjects.

Inspired by the promising WMn synthesis approach of Umapathy et al. (2021) and leveraging the contrast benefits of WMn-THOMAS, we introduce here a new pre-processing transform step, Histogram-based Polynomial Synthesis (HIPS), that enables robust and improved thalamic nuclei segmentation by synthesizing WMn-like images from T1w MRI using a simple polynomial function for intensity transformation. To test the performance of HIPS-THOMAS against T1w-THOMAS and CNN-based segmentation, quantitative performance metrics including Dice, volume errors, and inter-scanner/intra-scanner variability were used to assess performance across differing contrast (MPRAGE, SPGR, MP2RAGE), scanner manufacturers (Philips, GE, Siemens), and field strengths (1.5T, 3T, 7T).

## 2. Methods

### 2.1 Histogram-based polynomial synthesis (HIPS)

The main motivation of HIPS is to synthesize WMn-like images from T1w MRI using a simple polynomial function. The polynomial order and coefficients are estimated using a training set of 10 T1-MPRAGE and WMn-MPRAGE datasets acquired concurrently from the same subjects on a Philips 3T scanner. Data from the two image contrasts were then affinely registered using Advanced Normalization Tools (ANTS) (Avant, Tustison & Song, 2009) with WMn-MPRAGE as fixed image and T1-MPRAGE as moving image. Both images were first cropped, prior to registration to focus on the bilateral thalami and avoid the skull or fat/skin tissue. Note that this cropping was done automatically using a cropped mask template warped from template space as described in Su et al. (2019) and is already part of the THOMAS pipeline. Image histograms were computed using a central axial slice of the cropped images. To render the preprocessing step independent of scanner type and subjects, T1-MPRAGE images were then normalized using the WM signal and WMn-MPRAGE images were normalized using the CSF signal, both extracted from their respective image histograms, prior to fitting. To perform the normalization, the mode of the tissue of interest (WM in T1w, CSF in WMn-MPRAGE) was first computed from the histogram and the highest value shared by at least 1% of voxels exceeding the mode was used for the normalization (voxel value/normalizing value). This approach is like the WhiteStripe intensity normalization method (Shinohara et al., 2014) and is effective in removing the high intensity tail corresponding to artifacts and outlier intensities (Sun et al., 2015). Contrast transformation is essentially estimating a function that optimally maps the normalized T1w space to normalized WMn space. Polynomial functions of orders 1 (i.e. linear) to 4 (quartic) were tested using a curve fitting function (*scipy.optimize.curve\_fit*, python 3.9). For each order, a single aggregate function derived from the 10 training subjects was applied to a subset of T1w images (10 Philips 3T, 9 GE 3T, 10 Siemens 3T, and 8 Siemens 7T datasets) to generate the corresponding synthetic WMn images that were evaluated against the corresponding native WMn-MPRAGE data. The optimality of the transform function was assessed using two quantitative

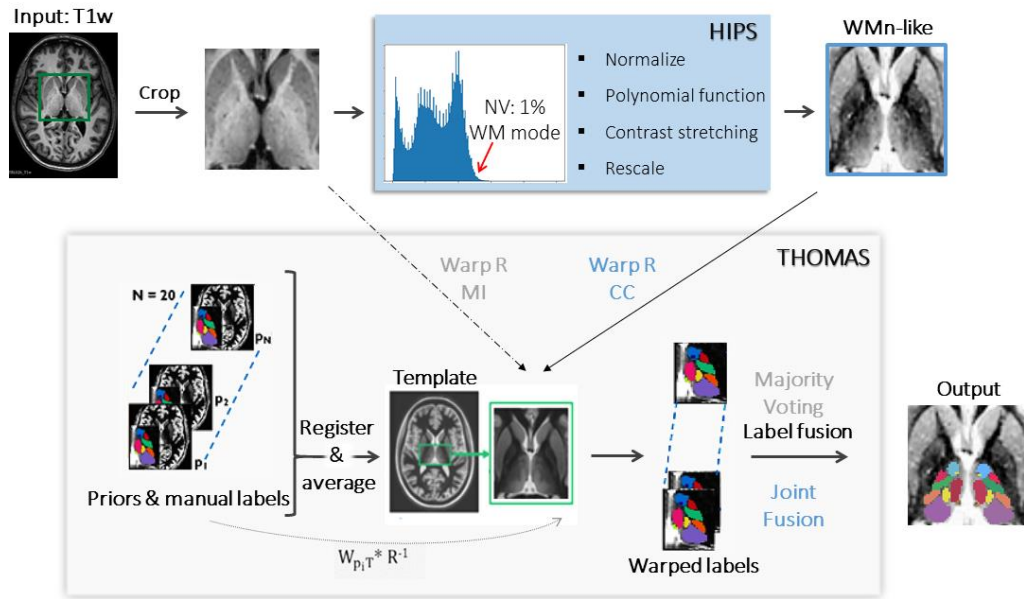


metrics- Structural Similarity Index (SSI) and Mean Square Error (MSE) and the best function was used for all subsequent processing.

## 2.2 The HIPS-THOMAS segmentation pipeline

HIPS-THOMAS is a variant of the THalamus Optimized Multi Atlas Segmentation (THOMAS) method (Su et al., 2019) and is shown in Figure 1. The input T1w image is first cropped to cover both thalami which removes outliers from the skull/subcutaneous fat. Following the cropping step, HIPS preprocessing is applied, which first consists of a normalization step described earlier, (“NV”, Fig. 1). The polynomial function is then applied followed by a contrast stretching step which rescales intensities within the percentiles range between 2 and 98 of the input image values and finally images are rescaled to the highest WM value computed previously with the histograms to get standardized intensity ranges. It results in a WMn-like version of the cropped input T1w image. This HIPS transformed cropped image is then nonlinearly registered (“Warp R”) to a cropped average brain template from 20 WMn-MPRAGE priors using the cross correlation metric (“CC” in blue, Fig. 1). This nonlinear warp is inverted (“ $R^{-1}$ ”, Fig.1) and combined with the 20 precomputed warps from priors to the template (“ $W_{piT}$ ”, Fig. 1) to put the 20 manual segmentation labels in the input space. These 20 labels were then combined using a joint label fusion algorithm to generate the output’s final parcellation. The segmented thalamic nuclei include Anteroventral (AV), Ventral anterior (VA), Ventral lateral anterior (VL<sub>a</sub>), Ventral lateral posterior (VL<sub>p</sub>), Ventral posterolateral (VPL), Pulvinar (Pul), Lateral geniculate (LGN), Medial geniculate (MGN), Centromedian (CM), Mediodorsal-Parafascicular (MD-Pf), and Habenular (Hb) in addition to the mammillothalamic tract (MTT).

The two notable differences between HIPS-THOMAS and T1w-THOMAS are the use of cross-correlation metric (“CC” in blue, Fig.1) for ANTs nonlinear registration in calculating the Warp R instead of mutual information metric (“MI” in grey, Fig.1) and the use of joint label fusion instead of majority voting in the final label fusion step.



**Fig. 1:** The proposed HIPS-THOMAS pipeline. HIPS pre-processing includes normalization of the cropped T1w input, application of a polynomial function to generate a WMn-like image, and a contrast stretching step followed by a rescaling step. The WMn-like cropped input is fed into the THOMAS pipeline as opposed to the original cropped T1w image. Note that for HIPS-THOMAS, the nonlinear warp R uses a cross-correlation metric (CC in blue) and the label fusion step uses joint label fusion (in blue) as opposed to mutual information metric (MI in grey) and majority voting (in grey).

### 2. 3 Convolutional Neural Network (CNN) based segmentation

In addition to T1w-THOMAS and HIPS-THOMAS, the dual CNN method of Umaphathy et al. (2021) was also used for segmentation of T1w data. This approach uses two cascaded 3D CNNs- WMn-MPRAGE-like images are first synthesized from T1w images using a contrast synthesis CNN and these images are then processed using another CNN to yield thalamic nuclei parcellations. The synthesis network was trained using patches from registered pairs of T1w and WMn-MPRAGE images acquired from the same subjects on 3T GE and Siemens scanners as described in Umaphathy et al (2021); the segmentation network was trained using WMn-THOMAS data.

## 2. 4 Datasets and evaluation metrics

The datasets used in the analysis comprised of 12 subjects acquired on a Siemens 3T scanner with T1w MPRAGE, 19 subjects acquired on a GE 3T scanner with 3D SPGR, and 18 subjects acquired on a Philips 3T scanner with T1w MPRAGE. WMn-MPRAGE acquired on each of these subjects were also available for comparisons. In addition, 8 datasets acquired on a Siemens 7T scanner using Magnetization-Prepared 2 Rapid Acquisition Gradient Echo (MP2RAGE) sequence were also analyzed. All experimental protocols including data acquisition were approved by institutional review board guidelines (University of Arizona for Siemens 3T data; Stanford and SRI International for GE 3T data; Comité de protection des personnes Ile-de-France IV for Philips 3T data; Commission cantonale d'éthique de la recherche sur l'être humain (CER-VD) for Siemens 7T data) and all data were acquired after obtaining prior written informed consent from the participants in accordance with the Declaration of Helsinki. Additional information about each sequence can be found in Supplementary Table 1. Segmentation performance of T1w-THOMAS, HIPS-THOMAS, and CNN were compared for different-

- i. T1w contrast (MPRAGE, SPGR, MP2RAGE)
- ii. scanner manufacturers (Siemens, GE, Philips)
- iii. field strengths (3T, 7T).

In the absence of “gold standard” manual segmentations, segmentations were compared to the “silver standard” WMn-THOMAS, using Dice coefficients and percentage volume error for each nucleus. Statistical significance was determined using t-tests with a Bonferroni correction (0.05/13) applied for multiple corrections.

## 2.5 Analysis of the Frequently Traveling Human Phantom (FTHP) MRI dataset

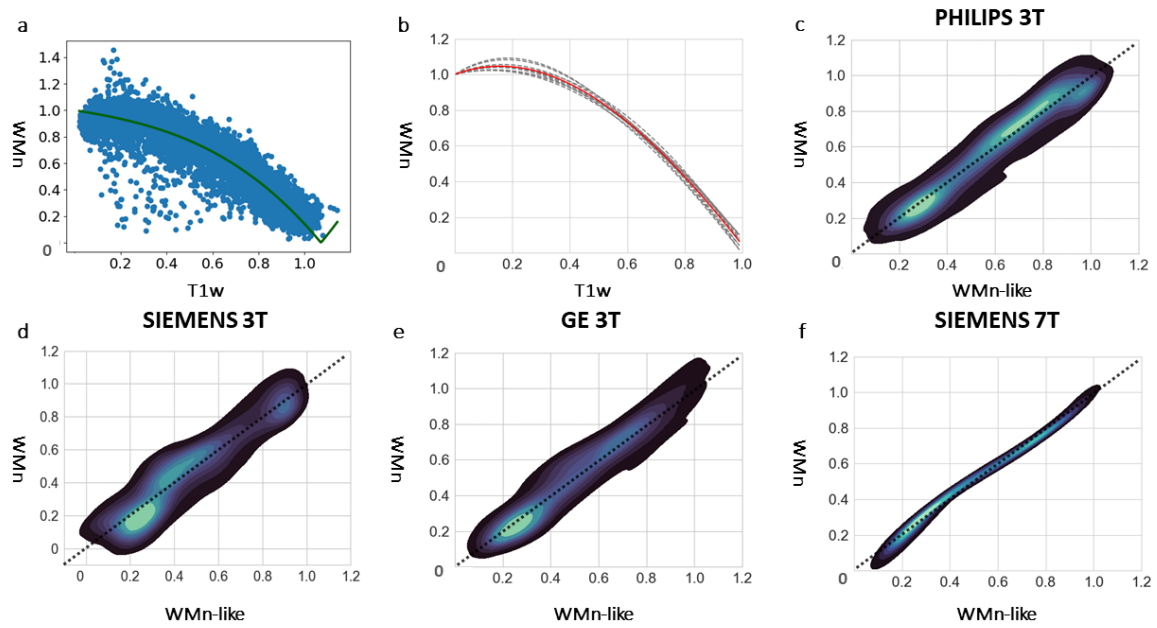
To assess the inter-scanner and intra-scanner variability of the different methods, we used a subset of the FTHP MRI dataset (Opfer et al., 2023). Briefly, this dataset comprises of T1w MRI from a single healthy male volunteer (age 49 years) scanned on 116 different MRI scanners. We used 3 scanner

manufacturers (Philips, Siemens, and GE), 2 field strengths (1.5T and 3T) and 4 different sites/scanner models for each manufacturer with 3 repeat scans at each site, resulting in a total of 72 scans. Intra-scanner variability was assessed from volume residuals by subtracting the mean volume of the 3 repeat scans acquired from the same scanner from individual volumes as in Opfer et al. (2023)., resulting in a vector of length 72. Inter-scanner variability was similarly assessed by first averaging the volumes from the three repeat scans and then computing residuals on the resulting vector of length 24.

### 3. Results

#### 3.1 HIPS transformation

For the polynomial orders considered, 2<sup>nd</sup> and 3<sup>rd</sup> order functions performed the best, based on SSI and MSE metrics. The only exception was Siemens 7T where the linear function performed the best. These results are summarized in Supplementary Table 2. The 3<sup>rd</sup> order polynomial equation was selected for use as we hypothesized more accuracy by considering the three tissues of interest (white matter, gray matter, cerebrospinal fluid). A plot of T1w vs. WMn-MPRAGE normalized image intensities (blue dots) for a representative Philips 3T subject is shown in Figure 2a with the best fit 3<sup>rd</sup> order curve shown in green. Individual best fit 3<sup>rd</sup> order curves for 10 Philips 3T subjects are shown in Figure 2b with the single curve aggregated from all 10 subjects shown overlaid in red. This curve was estimated as  $f(x) = 1 + 0.597*x - 2.0067*x^2 + 0.4529*x^3$ . Since this function was derived from normalized data, it was applicable across all subjects from different scanners and field strengths. Density plots between native cropped central slice WMn-MPRAGE and WMn-like synthesized images obtained through the application of the aggregated function on an example case from each of the 4 scanner types are shown in Figures 2c-f. The linearity of the density plots attests to the quality and robustness of the synthesis, with the Siemens 7T showing the most concordance (minimal dispersion from the unity line).

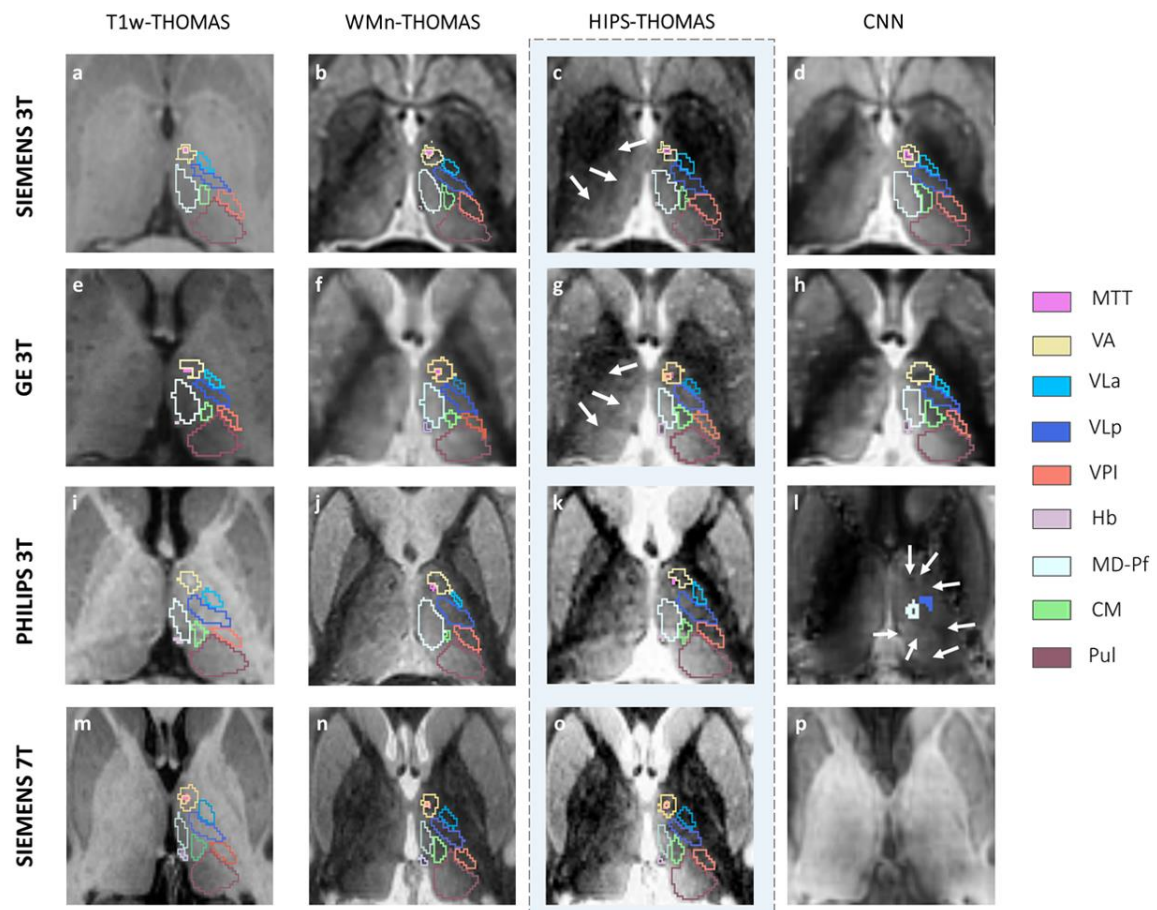


**Fig. 2:** HIPS formulation and performance. (a) Normalized intensity plot between WMn-MPRAGE and T1-MPRAGE data from a Philips 3T subject (blue dots) and a 3<sup>rd</sup> order polynomial fit (green line). (b) The 10 curves resulting from the individual fitting on 10 Philips data cases (dashed gray) and the resulting aggregated function (red line in a, b). Density plots between normalized WMn-MPRAGE and synthesized WMn-MPRAGE data using the same aggregated function on an example subject from Philips 3T (c), Siemens 3T (d), GE 3T (e), and Siemens 7T (f). The black dashed unity line represents perfect concordance between images.

### 3.2 Qualitative comparisons

Figure 3 shows acquired T1w and WMn-MPRAGE images as well as HIPS and CNN-synthesized WMn-like images for a Siemens 3T subject (a-d) and a GE 3T subject (e-h). The corresponding thalamic nuclei segmentations for the left side are also shown overlaid. Both the CNN and HIPS synthesized WMn images show improved intra-thalamic contrast (brighter signal in MD and pulvinar nuclei) and thalamic boundaries (white arrows) compared to T1w images. The CNN-synthesized images look less noisy with even better contrast compared to HIPS (d, h). T1w and WMn-MPRAGE images as well as HIPS and CNN-synthesized WMn-like images for a Philips 3T subject (i-l) and a Siemens 7T subject (m-

p) are also shown in Figure 3. Note that the CNN was trained only using GE and Siemens 3T data as described earlier and the Philips and 7T represent a different scan manufacturer and field strength, respectively, as a test of robustness. HIPS-synthesized images look very similar to WMn-MPRAGE images and produce segmentations comparable to those of WMn-THOMAS for both Philips 3T and Siemens 7T subjects. The CNN method failed on more than a third of the Philips cases (Fig. 3l) and on all Siemens 7T cases (Fig. 3p) due to failures in the synthesis step, leading to poor performance of the subsequent segmentation CNN.



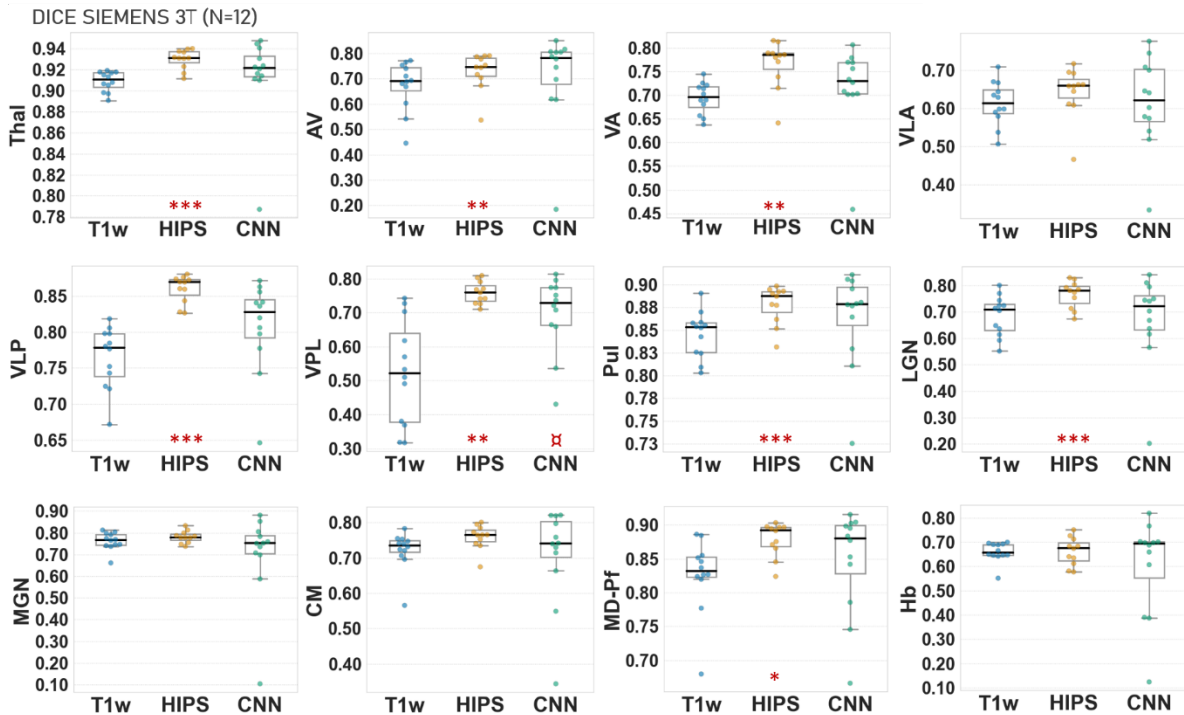
**Fig. 3:** An axial slice from acquired T1w and WMn-MPRAGE as well as HIPS and CNN-synthesized WMn images for a Siemens 3T MPRAGE (a-d), GE 3T SPGR (e-h), Philips 3T MPRAGE (i-l) or Siemens 7T MP2RAGE (m-p) subject with the corresponding nuclei segmentations overlaid on the left thalamus. Note the improved intrathalamic contrast and thalamic boundaries (white arrows on c and g) in the

synthesized WMn-like images produced by HIPS-THOMAS compared to the native T1w images. The failure of CNN synthesis can clearly be seen panels l and p along with the failed or missing (white arrows in l) segmentations.

### 3. 3 Quantitative assessments

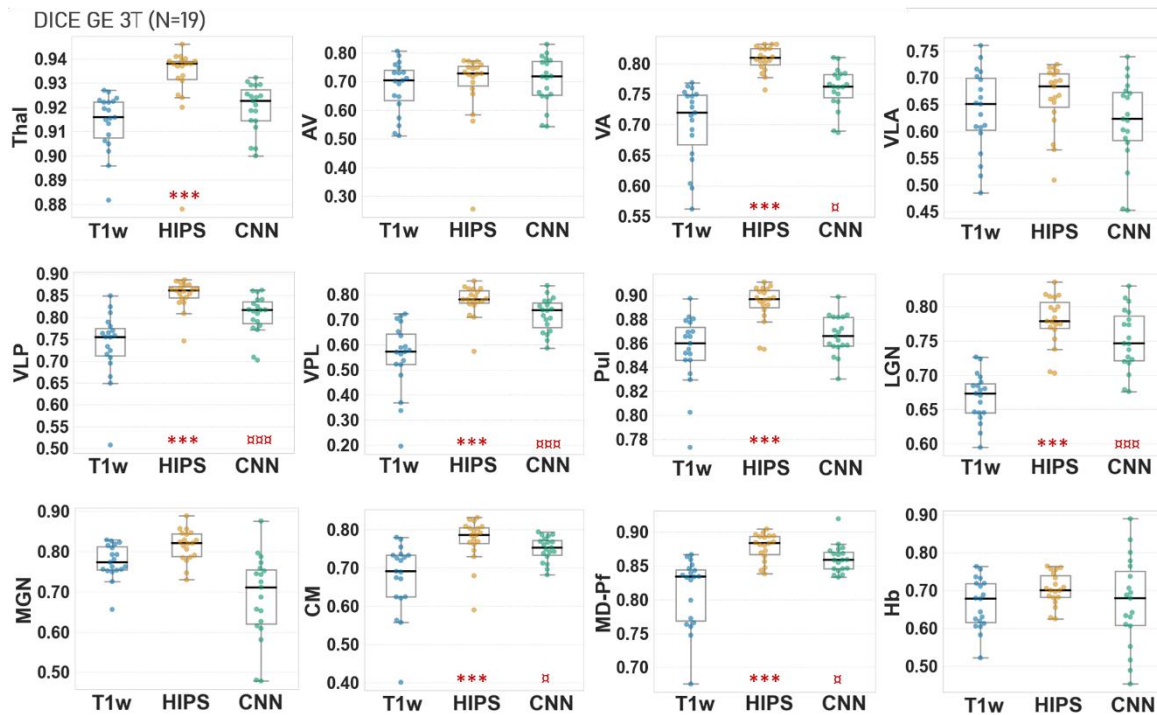
For quantitative comparisons, mean Dice coefficients from Siemens 3T MPRAGE (n=12) and GE 3T SPGR (n=19) datasets were compared between T1w-THOMAS, HIPS-THOMAS, and CNN using THOMAS segmentation from WMn-MPRAGE as a silver-standard and the % improvement over T1w-THOMAS was computed. HIPS-THOMAS showed significantly improved Dice compared to T1w-THOMAS for 7/11 nuclei and the whole thalamus on both Siemens 3T and GE data (Figures 4-5). The raw data are tabulated in Supplementary Tables 3 and 4 for the Siemens 3T MPRAGE and the GE 3T SPGR datasets respectively. For HIPS-THOMAS, more than 10% increase in Dice for the VA, VLP, VPL, and LGN nuclei was observed on Siemens 3T MPRAGE and more than 15% increase in Dice for the VA, VLP, VPL, LGN, and CM nuclei was observed on GE 3T SPGR data. By contrast, the CNN performed better on 1/11 nuclei on Siemens 3T data and 6/11 on GE 3T data (“x” in Fig. 4 and Fig. 5) with decreased Dice for several nuclei significant only for the MTT (Supp. Tab. 4). Notably, HIPS-THOMAS outperformed the CNN, with a higher mean and lower standard deviation, for many nuclei, especially the VLP and the MTT for both GE and Siemens data but also the VA, VPL, Pul, MGN, and MD on GE 3T data. Like the performance on GE and Siemens 3T, HIPS-THOMAS significantly improved Dice compared to T1w-THOMAS for 8/11 nuclei and the whole thalamus on Philips 3T data (Fig. 6) and 9/11 nuclei and the whole thalamus on Siemens 7T data (Fig. 7) with the corresponding data tabulated in Supplementary Table 5. More than 15% increase in Dice for the VA, VLP, MD nuclei was observed on Philips 3T data and more than 30% increase for the AV, VA, VL<sub>a</sub>, VLP, VPL, CM, MD nuclei on the Siemens 7T data. Only the Hb nucleus showed decreased Dice using HIPS-THOMAS compared to T1w-THOMAS on SIEMENS 7T (not statistically significant). Subjects who were scanned on 7T were also

scanned on 3T and this data is reported in Supplementary Table 5. HIPS-THOMAS at 7T is substantially better (>15%) than at 3T on the VLa, LGN, and CM nuclei as well as the MTT.

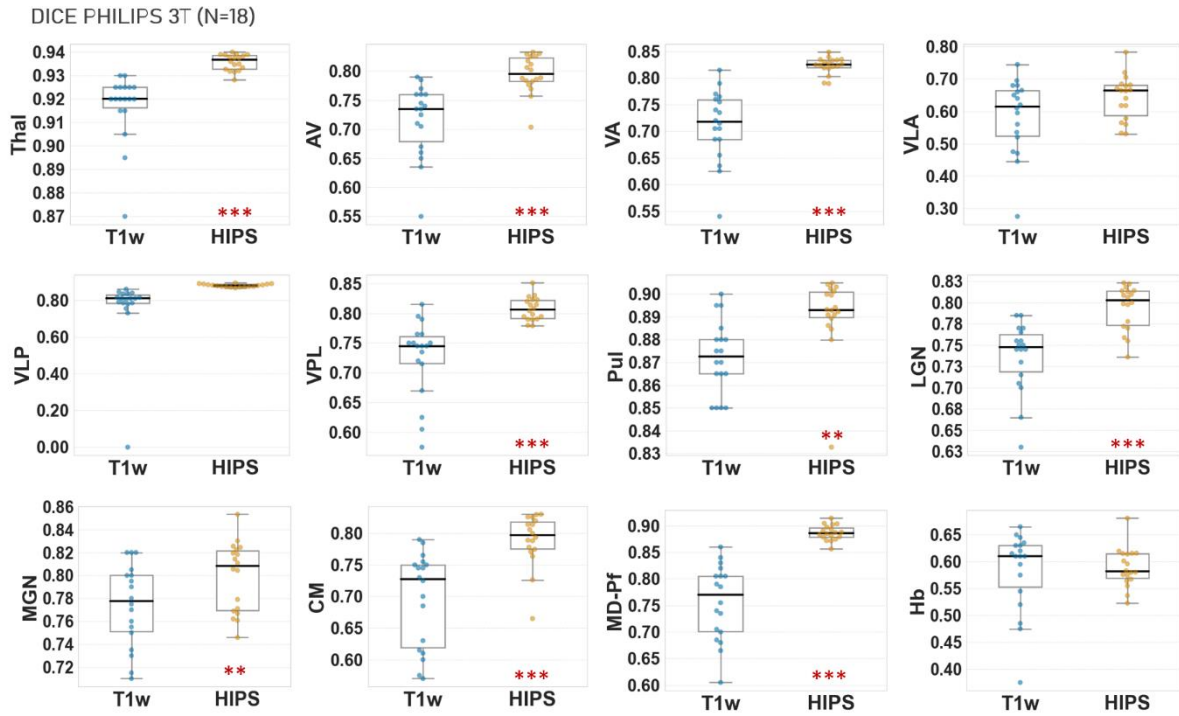


**Fig.4:** Box plots of T1w-THOMAS (T1w), HIPS-THOMAS (HIPS), and CNN segmentation’s Dice coefficients (compared against WMn-THOMAS segmentations) for 3T Siemens MPRAGE data (n=12). HIPS-THOMAS significantly improves Dice in whole thalamus and 7 nuclei compared to 1 nucleus for CNN. T-test between the Dice coefficients of the T1w-THOMAS vs. HIPS-THOMAS: P-values \*<0.00385 (Bonferonni correction for multiple comparisons, 0.05/13) \*\*<0.001 \*\*\*<0.0001. T1w-THOMAS vs. CNN:  $\alpha$  p<0.00385 (Bonferonni correction for multiple comparisons, 0.05/13),  $\beta$  p<0.001,  $\beta\beta$  p<0.0001.

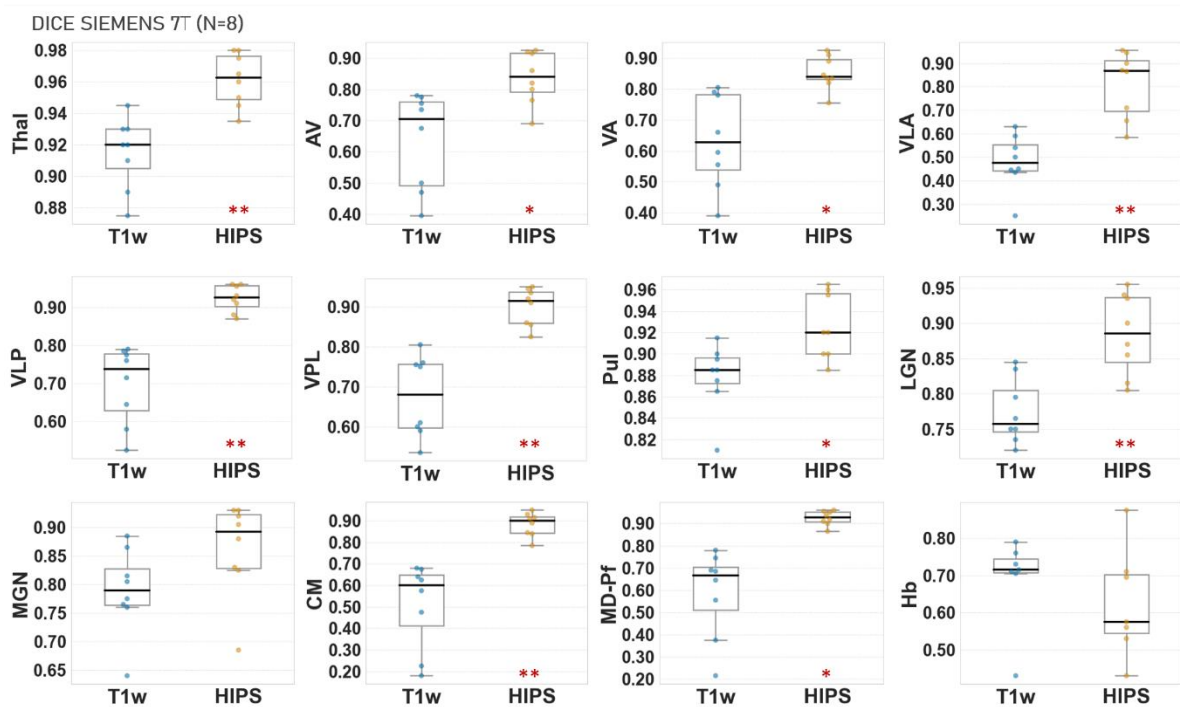




**Fig.5:** Box plots of T1w-THOMAS (T1w), HIPS-THOMAS (HIPS), and CNN segmentation’s Dice coefficients (compared against WMn-THOMAS segmentations) for GE 3T SPGR data (n=19). HIPS-THOMAS significantly improves Dice in whole thalamus and 7 nuclei compared to 6 nuclei for CNN. T-test between the Dice coefficients of the T1w-THOMAS vs. HIPS-THOMAS: P-values  $* < 0.00385$  (Bonferonni correction for multiple comparisons,  $0.05/13$ )  $** < 0.001$   $*** < 0.0001$ . T1w-THOMAS vs. CNN:  $\times p < 0.00385$  (Bonferonni correction for multiple comparisons,  $0.05/13$ ),  $\times \times p < 0.001$ ,  $\times \times \times p < 0.0001$ .



**Fig.6:** Box plots of T1w-THOMAS (T1w) and HIPS-THOMAS (HIPS) Dice coefficients (compared against WMn-THOMAS segmentations) for Philips 3T MPRAGE data (n=18). HIPS-THOMAS significantly improves Dice in whole thalamus and 8 nuclei. T-test between the DICE coefficients of the T1w-THOMAS VS HIPS-THOMAS: P-values  $* < 0.00385$  (Bonferonni correction for multiple comparisons,  $0.05/13$ )  $** < 0.001$   $*** < 0.0001$ .



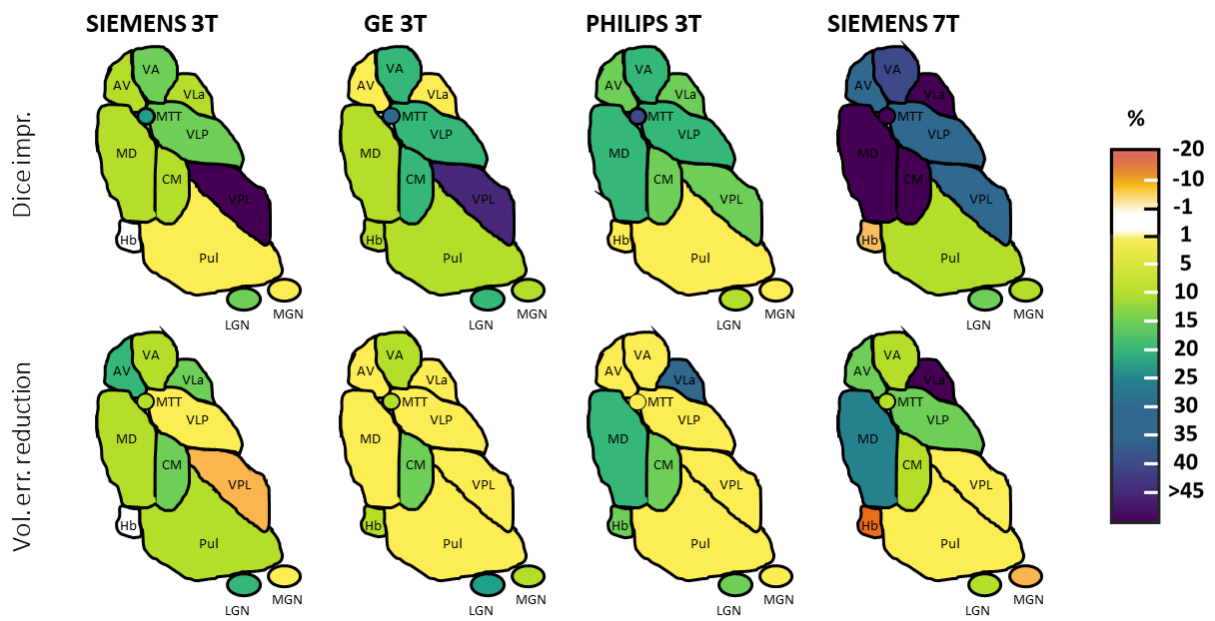
**Fig.7:** Box plots of T1w-THOMAS (T1w) and HIPS-THOMAS (HIPS) Dice coefficients (compared against WMn-THOMAS segmentations) for Siemens 7T MP2RAGE data (n=8). HIPS-THOMAS improves Dice in whole thalamus and 9 nuclei. T-test between the Dice coefficients of the T1w-THOMAS VS HIPS-THOMAS: P-values \* $<0.00385$  (Bonferonni correction for multiple comparisons,  $0.05/13$ ) \*\* $<0.001$  \*\*\* $<0.0001$ .

Mean volume errors (expressed as percentage) of T1w-THOMAS, HIPS-THOMAS, and CNN segmentations compared to WMn-THOMAS segmentation are shown in Figure 8 for Siemens 3T (a), GE 3T (b), Philips 3T (c), and Siemens 7T (d) datasets. HIPS-THOMAS had the **lowest** error in 9 nuclei and the MTT and highest error in 2 nuclei for Siemens 3T data. A similar trend was also observed for the GE 3T data. T1w-THOMAS displayed the **highest** errors in 8 nuclei on Siemens 3T data and 7 nuclei on GE 3T data. CNN performance was either comparable to or slightly worse than HIPS except in the MTT and the AV nucleus on Siemens data, where it was substantially worse (higher error) and the AV nucleus on GE data, where it was substantially better (lower error). HIPS-THOMAS had lower mean errors on all nuclei for Philips 3T data and all except the Hb nucleus for Siemens 7T data.



**Fig.8:** Mean volume error (%) of T1w-THOMAS, HIPS-THOMAS, and CNN segmentations, compared to WMn-THOMAS segmentations for Siemens 3T MPRAGE (a, n=12) and GE 3T SPGR (b, n=19) Philips 3T MPRAGE 3T (c, n=18) and Siemens 7T MP2RAGE (d, n=8) data. The general trend of HIPS < CNN < T1w THOMAS was observed for most nuclei at 3T. HIPS-THOMAS errors were lower than T1w-THOMAS for all nuclei but Hb for Siemens 7T data.

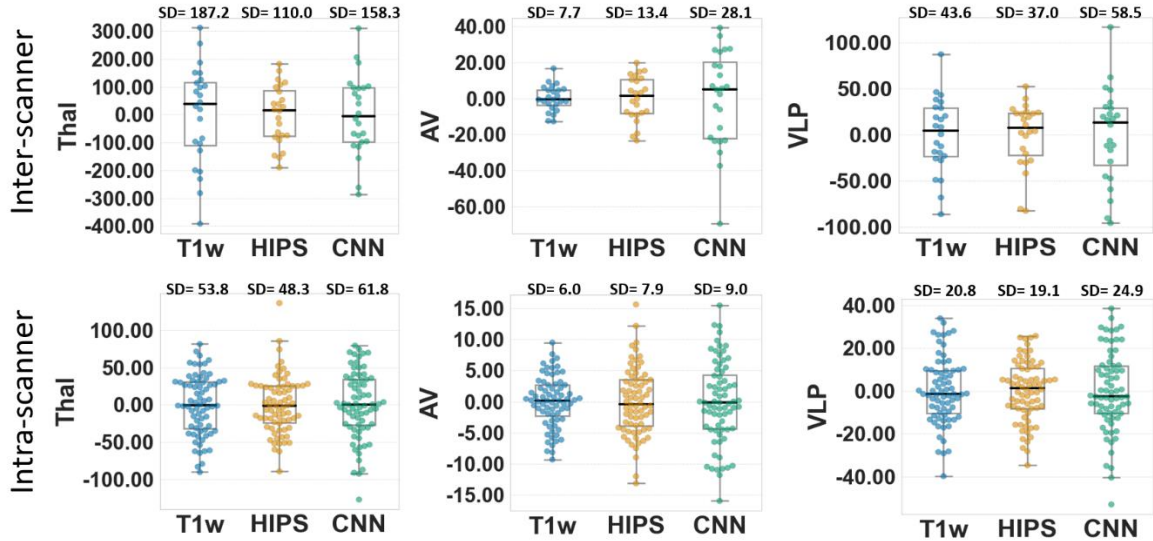
Figure 9 visually summarizes the improvements in Dice (%) and reduction in volume error % for HIPS-THOMAS compared to T1w-THOMAS for the 4 scanners (GE 3T, Siemens 3T, Philips 3T, and Siemens 7T). While 7T HIPS-THOMAS displayed the largest increase in Dice coefficients and reduction in volume errors, 3T showed at least 15% increase in Dice in ventral nuclei such as the VA and VPL across all scanners. Volume errors were also reduced for HIPS-THOMAS for all the scanners except for the VPL and Hb nuclei on GE 3T data and the Hb nucleus on Siemens 7T data (white regions).



**Fig.9:** Graphical summary of improvement of mean Dice coefficients (%) and reduction of % volume error for each thalamic nucleus on Siemens 3T, GE 3T, PHILIPS 3T and Siemens 7T data using HIPS-THOMAS compared to T1w-THOMAS. Labels: cf. Method 2.2.

### 3. 4 Inter- and intra-scanner variability

Figure 10 shows inter-scanner and intra-scanner variability for the whole thalamus (Thal), a small nucleus (AV) and a large nucleus (VLP). The results for all the thalamic nuclei are reported in Supplementary Table 6. The standard deviation of the residuals is reported on the top of each panel. HIPS-THOMAS had the least inter-scanner variability for the whole thalamus and 8/11 nuclei. It also had the least intra-scanner variability for the whole thalamus and 4/11 nuclei. In contrast, T1w-THOMAS had the least inter-scanner variability for 4/11 nuclei and least intra-scanner variability for 8/11 nuclei. CNN had the worst inter- and intra-scanner variability for the whole thalamus and all nuclei.



**Fig.10:** Inter-scanner and intra-scanner variability for the Frequently Traveling Human Phantom MRI dataset (in units of mm<sup>3</sup>) for T1w-THOMAS, HIPS-THOMAS, and CNN segmentation of the whole thalamus (Thal), AV and VLP nuclei. For this analysis, 24 different scanners covering 3 manufacturers, 2 field strengths and 4 sites each with 3 repeat scans per site were used, resulting in a dataset of 72 scans. Standard deviations (SD) for each method is shown above each plot.

#### 4. Discussion

We have significantly improved the THOMAS pipeline which was developed for 7T WMn-MPRAGE data to allow thalamic nuclei segmentation from standard 3T T1w images, by employing a computationally efficient polynomial synthesis transform to generate WMn-like images prior to segmentation. Robustness of segmentations using HIPS-THOMAS and a single cubic function estimated from aggregating 10 3T datasets was demonstrated from data across differing T1 image contrasts (i.e. SPGR, MPRAGE, MP2RAGE), scanner manufacturers (i.e. Siemens, GE, Philips), and field strengths (i.e. 3T, 7T). HIPS-THOMAS had higher Dice coefficients, lower volume errors, and lower inter-scanner variability compared to T1w-THOMAS and CNN for most of the nuclei.

Segmentations of T1w data using HIPS-THOMAS are closer than T1w-THOMAS to THOMAS segmentations using WMn-MPRAGE images, resulting in improved volume accuracy and Dice coefficients for most of the nuclei. Dice was most improved for ventral nuclei (VA, VLP, VPL) where the improved contrast in T1w images better delineates external boundaries. The HIPS-THOMAS Dice improvement reaches its highest performance at 7T, likely due to the more accurate synthesis of WMn-like images at 7T using MP2RAGE T1 maps. Only the Hb nucleus has a decreased Dice index using HIPS-THOMAS on Siemens 7T data compared to the use of T1w-THOMAS (although not statistically significant), which may be explained by a higher variability of segmentations for this nucleus and a small effective in this dataset. In contrast, T1w-THOMAS showed much worse sensitivity for the left VPL nucleus ( $p=0.011$ ) and a potentially spurious atrophy of the right VPL nucleus. These results are analogous to the results of Umapathy et al. (2021) where the synthesis-segmentation CNN displayed better accuracy compared to a segmentation network trained directly on T1w images, which showed spurious VPL atrophy. Results from the FTHP MRI dataset indicate that HIPS-THOMAS showed the least inter-scanner variability for the whole-thalamus and most nuclei. T1w-THOMAS, however, had lower intra-scanner variability for more nuclei than HIPS-THOMAS. This could be because the FTHP dataset included 1.5T MRI data which may not be optimal for HIPS or CNN. The standard deviation for intra-scanner variability for the whole thalamus using HIPS-THOMAS is comparable to the values reported in Opfer et al. (2023) using their custom CNN method (which only segments the whole thalamus and not nuclei) and FastSurfer (48 mm<sup>3</sup> HIPS vs. 40 mm<sup>3</sup> Opfer CNN, 45 mm<sup>3</sup> FastSurfer). The standard deviation for inter-scanner variability for the whole thalamus using HIPS-THOMAS is significantly lower than the values reported in Opfer et al. (2023) (110 mm<sup>3</sup> HIPS vs. 140 mm<sup>3</sup> Opfer CNN, 310 mm<sup>3</sup> FastSurfer).

This work used data from 3T Philips subjects to compute an aggregate 3<sup>rd</sup> order function that was used for all subjects. The use of just Philips data is perhaps a limitation and linked to the non-availability of concurrent WMn-T1 data from other scanners at the commencement of this project. However, to

confirm its validity across data from all scanners, on a small subset, we compared this function to a function derived from mixed data (10 Philips 3T MPRAGE, 9 GE 3T SPGR, 10 Siemens 3T MPRAGE and 8 Siemens 7T MP2RAGE cases) and found no significant differences in SSI or MSE (Supp. Tab. 7). This is further confirmed by the similarity of transformations between resulting WMn-like images after the application of the Philips equation vs. mixed equation on T1w-MRI (Supp. Fig. 1). At 7T, the MP2RAGE sequence produces significantly higher contrast images than MPRAGE, resulting in synthesized WMn-like images that very closely match the WMn-MPRAGE images as seen by the near perfect linear density plot with minimal spread in Figure 2F. The 3<sup>rd</sup> order polynomial performed optimally at 3T. However, at 7T, the linear function performed better, and this is reflected by the almost perfect unity line of the joint distribution plots in Figure 2F as well as better metrics and synthesized WMn-like images (Supp. Tab. 7 and Supp. Fig. 1). Given that 3T is the most common field strength used for neuroimaging and that the Dice improvements using the 3<sup>rd</sup> order function was still very significant at 7T, we made the 3<sup>rd</sup> order function as default in HIPS-THOMAS. (Note that HIPS-THOMAS can be run with a user-specified polynomial order and 7T data should be run with linear mode to get the best results). Future work will examine if use of functions optimized for each scanner and field strength is more beneficial as the exact nature of the curve depends on the parameters of the T1-MPRAGE sequence used, even though our normalization process helps reduce some of that variability. The significant contribution of HIPS to the THOMAS pipeline is to create images with similar contrast profiles as WMn-MPRAGE images, allowing the use of CC metric for nonlinear registration to the WMn template, which is more accurate than the MI metric used in T1w-THOMAS (Andronache et al., 2008). The joint fusion algorithm used in HIPS-THOMAS (cf. majority voting in T1w-THOMAS), also likely contributed to increased label accuracy (Bernstein et al., 2021; Pfefferbaum et al., 2023). HIPS is computationally efficient, does not add much complexity to the image analysis pipeline of THOMAS, and does not require separate training for the different scenarios of scanner manufacturers and field strengths, as required by CNN-based methods.



The CNN-synthesized images have less noise and slightly enhanced contrast compared to HIPS images (Fig. 3) due to denoising inherent in the synthesis CNN. In contrast, the use of polynomial functions in HIPS could result in noise amplification. Future work could address this using denoising methods. CNN performs comparably to HIPS-THOMAS on sequences on which it was trained (i.e. Siemens 3T MPRAGE) or on similar contrast images (SPGR), but HIPS-THOMAS is considerably better than the CNN for Philips 3T and Siemens 7T data, where the CNN failed on most cases due to lack of adequate training. While HIPS-THOMAS can yield higher Dice coefficients, it may paradoxically have higher volume errors than the CNN for certain nuclei on Siemens 3T (e.g. VPL) and GE 3T data (e.g. AV, LGN), as illustrated in Figure 4. This discrepancy may be due to the denoising present in CNNs, which allows for more precise delimitation. In summary, HIPS-THOMAS is more flexible and generalizable, as it can be applied easily to T1w images from different scanners without requiring training, a big advantage considering public databases like ADNI and OASIS contain data from a mix of manufacturers and field strengths.

Our work had some limitations. The T1w segmentations were evaluated against segmentations from THOMAS applied to WMn-MPRAGE images, which is not ideal. THOMAS has been thoroughly validated against manual segmentation at 3T (Su et al, ISMRM abstract) (Su et al., 2016) and 7T (Su et al., 2019) and was adopted as a “silver” standard, substituting for the “gold” standard manual segmentation, which is very time consuming and requires specific domain expertise. The performance of the CNN method could also be enhanced by training the synthesis and segmentation networks with data from Philips and 7T scanners but was beyond the scope of this work. In the future, more sophisticated image and histogram normalization like that proposed in Nyul et al. (1999) or more complex exemplar-based synthesis approaches such as MIMECS (MR image example-based contrast synthesis) (Roy et al., 2011), could replace the simpler 3<sup>rd</sup> order polynomial approach taken here, which could further improve HIPS performance and reduce noise amplification. This could also be useful for pipelines for analyzing whole

brain images (as opposed to cropped images like in our current pipeline) where artifacts from scalp and other sources can impair HIPS performance. HIPS showed Dice improvements in several nuclei, but the Hb and VL<sub>a</sub> nuclei consistently showed < 0.7 Dice across all 3T scanners. While the poor habenula Dice could be attributed to its small size, the reasons for suboptimal VL<sub>a</sub> Dice compared to other ventral nuclei like the VA warrants further investigation.

## 5. Conclusion

WMn-like images synthesized using the computationally efficient HIPS significantly improved the robustness as well as the accuracy of THOMAS compared to direct THOMAS or CNN-based methods for segmentation of T1w data.

## 6. Data availability statement

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

## 7. Code availability

The code used in the current study for generating the polynomial fits as well as the HIPS transformation and the integrated HIPS-THOMAS pipeline is available at <https://github.com/thalamicseg/hipsthomasdocketer>.

## 8. References

[1] Morel, A., Magnin, M. & Jeanmonod, D. Multiarchitectonic and stereotactic atlas of the human thalamus. *J. Comp. Neurol.* 387:588–630 (1997).

[2] Sherman, S. M. The thalamus is more than just a relay. *Current opinion in neurobiology*, 17(4), 417-422 (2007).

- [3] Zahr, N. M., Sullivan, E. V., Pohl, K. M., Pfefferbaum, A., & Saranathan, M. Sensitivity of ventrolateral posterior thalamic nucleus to back pain in alcoholism and CD4 nadir in HIV. *Human brain mapping*, 41(5), 1351-1361 (2020).
- [4] Andreasen, N. C. The evolving concept of schizophrenia: from Kraepelin to the present and future. *Schizophrenia research*, 28(2-3), 105-109 (1997).
- [5] Braak, H., & Braak, E. Demonstration of amyloid deposits and neurofibrillary changes in whole brain sections. *Brain pathology*, 1(3), 213-216 (1991).
- [6] Gustin, S. M., Peck, C. C., Wilcox, S. L., Nash, P. G., Murray, G. M., & Henderson, L. A. Different pain, different brain: thalamic anatomy in neuropathic and non-neuropathic chronic pain syndromes. *Journal of Neuroscience*, 31(16), 5956-5964 (2011).
- [7] Fisher, R., Salanova, V., Witt, T., Worth, R., Henry, T., Gross, R., ... & SANTE Study Group. Electrical stimulation of the anterior nucleus of thalamus for treatment of refractory epilepsy. *Epilepsia*, 51(5), 899-908 (2010).
- [8] Danet, L., Barbeau, E. J., Eustache, P., Planton, M., Raposo, N., Sibon, I., ... & Pariente, J. Thalamic amnesia after infarct: the role of the mammillothalamic tract and mediodorsal nucleus. *Neurology*, 85(24), 2107-2115 (2015).
- [9] Koller, W. C., Lyons, K. E., Wilkinson, S. B., Troster, A. I., & Pahwa, R. Long-term safety and efficacy of unilateral deep brain stimulation of the thalamus in essential tremor. *Movement disorders: official journal of the Movement Disorder Society*, 16(3), 464-468 (2001).
- [10] Owen, S. L., Green, A. L., Stein, J. F., & Aziz, T. Z. Deep brain stimulation for the alleviation of post-stroke neuropathic pain. *Pain*, 120(1-2), 202-206 (2006).
- [11] Benabid, A. L., Pollak, P., Louveau, A., Henry, S., & De Rougemont, J. Combined (thalamotomy and stimulation) stereotactic surgery of the VIM thalamic nucleus for bilateral Parkinson disease. *Stereotactic and functional neurosurgery*, 50(1-6), 344-346 (1987).
- [12] Cinalli, G., Aguirre, D. T., Mirone, G., Ruggiero, C., Cascone, D., Quaglietta, L., ... & Spennato, P. Surgical treatment of thalamic tumors in children. *Journal of Neurosurgery: Pediatrics*, 21(3), 247-257 (2018).
- [13] Schaltenbrand, G. & Warren, W. Atlas for Stereotaxy of the Human Brain (ed 2. Stuttgart), Thieme, (1977).
- [14] Sanborn, M. R., Danish, S. F., Ranalli, N. J., Grady, M. S., Jaggi, J. L., & Baltuch, G. H. Thalamic deep brain stimulation for midbrain tremor secondary to cystic degeneration of the brainstem. *Stereotactic and Functional Neurosurgery*, 87(2), 128-133 (2009).
- [15] Jakab, A., Blanc, R., Berényi, E. L., & Székely, G. (2012). Generation of individualized thalamus target maps by using statistical shape models and thalamocortical tractography. *American Journal of Neuroradiology*, 33(11), 2110-2116.
- [16] Behrens, T. E., Johansen-Berg, H., Woolrich, M. W., Smith, S. M., Wheeler-Kingshott, C. A. M., Boulby, P. A., ... & Matthews, P. M. Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. *Nature neuroscience*, 6(7), 750-757 (2003).
- [17] Mang, S. C., Busza, A., Reiterer, S., Grodd, W., & Klose, A. U. Thalamus segmentation based on the local diffusion direction: a group study. *Magnetic resonance in medicine*, 67(1), 118-126 (2012).

- [18] Battistella, G., Najdenovska, E., Maeder, P., Ghazaleh, N., Daducci, A., Thiran, J. P., ... & Fornari, E. Robust thalamic nuclei segmentation method based on local diffusion magnetic resonance properties. *Brain Structure and Function*, 222, 2203-2216 (2017).
- [19] Zhang, D., Snyder, A. Z., Fox, M. D., Sansbury, M. W., Shimony, J. S., & Raichle, M. E. Intrinsic functional relations between human cerebral cortex and thalamus. *Journal of neurophysiology*, 100(4), 1740-1748 (2008).
- [20] Hale, J. R., Mayhew, S. D., Mullinger, K. J., Wilson, R. S., Arvanitis, T. N., Francis, S. T., & Bagshaw, A. P. Comparison of functional thalamic segmentation from seed-based analysis and ICA. *Neuroimage*, 114, 448-465 (2015).
- [21] Kumar, V. J., van Oort, E., Scheffler, K., Beckmann, C. F., & Grodd, W. Functional anatomy of the human thalamus at rest. *Neuroimage*, 147, 678-691 (2017).
- [22] Van Oort, E. S., Mennes, M., Schröder, T. N., Kumar, V. J., Jimenez, N. I. Z., Grodd, W., ... & Beckmann, C. F. Functional parcellation using time courses of instantaneous connectivity. *NeuroImage*, 170, 31-40 (2018).
- [23] Liu, Y., D'Haese, P. F., Newton, A. T., & Dawant, B. M. 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, 114-128 (2020).
- [24] Deoni, S. C., Josseau, M. J., Rutt, B. K., & Peters, T. M. Visualization of thalamic nuclei on high resolution, multi-averaged T1 and T2 maps acquired at 1.5 T. *Human brain mapping*, 25(3), 353-359 (2005).
- [25] Iglesias, J. E., Insausti, R., Lerma-Usabiaga, G., Bocchetta, M., Van Leemput, K., Greve, D. N., ... & Alzheimer's Disease Neuroimaging Initiative. A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *Neuroimage*, 183, 314-326 (2018).
- [26] Tourdias, T., Saranathan, M., Levesque, I. R., Su, J., & Rutt, B. K. Visualization of intra-thalamic nuclei with optimized white-matter-nulled MPRAGE at 7 T. *Neuroimage*, 84, 534-545 (2014).
- [27] Magnotta, V. A., Gold, S., Andreasen, N. C., Ehrhardt, J. C., & Yuh, W. T. Visualization of subthalamic nuclei with cortex attenuated inversion recovery MR imaging. *Neuroimage*, 11(4), 341-346 (2000).
- [28] Su, J. H., Thomas, F. T., Kasoff, W. S., Tourdias, T., Choi, E. Y., Rutt, B. K., & Saranathan, M. Thalamus Optimized Multi Atlas Segmentation (THOMAS): fast, fully automated segmentation of thalamic nuclei from structural MRI. *Neuroimage*, 194, 272-282 (2019).
- [29] Su, J. H. et al. Improved Vim targeting for focused ultrasound ablation treatment of essential tremor: A probabilistic and patient-specific approach. *Human Brain Mapping*, 41(17), 4769-4788 (2020). Improved Vim targeting for focused ultrasound ablation treatment of essential tremor: A probabilistic and patient-specific approach. *Human Brain Mapping*, 41(17), 4769-4788 (2020).
- [30] Bernstein, A. S., Rapcsak, S. Z., Hornberger, M., Saranathan, M., & Alzheimer's Disease Neuroimaging Initiative. Structural changes in thalamic nuclei across prodromal and clinical Alzheimer's disease. *Journal of Alzheimer's Disease*, 82(1), 361-371 (2021).
- [31] Pfefferbaum, A., Sullivan, E. V., Zahr, N. M., Pohl, K. M., & Saranathan, M. Multi-atlas thalamic nuclei segmentation on standard T1-weighted MRI with application to normal aging. *Human Brain Mapping*, 44(2), 612-628 (2023).

- [32] Umapathy, L., Keerthivasan, M. B., Zahr, N. M., Bilgin, A., & Saranathan, M. Convolutional Neural Network Based Frameworks for Fast Automatic Segmentation of Thalamic Nuclei from Native and Synthesized Contrast Structural MRI. *Neuroinformatics*, 1-14 (2021).
- [33] Avants, B. B., Tustison, N., & Song, G. Advanced normalization tools (ANTS). *Insight j*, 2(365), 1-3. (2009).
- [34] Shinohara, R. T., Sweeney, E. M., Goldsmith, J., Shiee, N., Mateen, F. J., Calabresi, P. A., ... & Alzheimer's Disease Neuroimaging Initiative. Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical*, 6, 9-19 (2014).
- [35] Sun, X., Shi, L., Luo, Y., Yang, W., Li, H., Liang, P., ... & Wang, D. Histogram-based normalization technique on human brain magnetic resonance images from different acquisitions. *Biomedical engineering online*, 14(1), 1-17 (2015).
- [36] Opfer, R., Krüger, J., Spies, L., Ostwaldt, A. C., Kitzler, H. H., Schippling, S., & Buchert, R. Automatic segmentation of the thalamus using a massively trained 3D convolutional neural network: higher sensitivity for the detection of reduced thalamus volume by improved inter-scanner stability. *European Radiology*, 33(3), 1852-1861 (2023).
- [37] Andronache, A., von Siebenthal, M., Székely, G., & Cattin, P. Non-rigid registration of multi-modal images using both mutual information and cross-correlation. *Medical image analysis*, 12(1), 3-15 (2008).
- [38] Su, J., Tourdias, T., Saranathan, M., Ghanouni, P., Rutt, B. THOMAS: Thalamus Optimized Multi-Atlas Segmentation at 3T. Proceedings of the 24th annual meeting of the ISMRM p4328 (2016).
- [39] Nyúl, L. G., & Udupa, J. K. On standardizing the MR image intensity scale. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 42(6), 1072-1081 (1999).
- [40] Roy, S., Carass, A., & Prince, J. A compressed sensing approach for MR tissue contrast synthesis. In Information Processing in Medical Imaging: 22nd International Conference, IPMI 2011, Kloster Irsee, Germany, July 3-8, 2011. Proceedings 22 (pp. 371-383). Springer Berlin Heidelberg (2011).

## 9. Acknowledgements

We wish to acknowledge Germain Arribarat for his support and his time during the beginning of this project, forming J.P.V to image processing. This work was supported by the University of Toulouse 3 Paul Sabatier via a PhD student mobility grant obtained by J.P.V and the Swiss National Science Foundation. M.S would like to acknowledge funding from the National Institute of Biomedical Imaging and Bioengineering (7R01EB032674).

## 10. Author contributions

J.P.V. and M.S. conceived and developed the presented method. J.P.V. performed the computations and analysis under the supervision of M.S. J.P.V. wrote the manuscript with support from L.D., E.J.B., and mainly with M.S.

L.D., E.J.B., P.P., J.P., M.B.C., N.M.Z. acquired and shared MRI data from their previous respective studies and helped with the study design.

## 11. Additional information

Competing interests' statement: The authors have no conflicts of interest relevant to this manuscript to disclose.

## 12. Supplementary files

		Philips	GE	SIEMENS (1)	SIEMENS (2)
<i>Data source</i>		[1]	[2]	[3]	[4]
<i>N</i>		18	19	12	9
<i>Field strength</i>		3T	3T	3T	7T
<i>Sequence</i>		MPRAGE	SPGR	MPRAGE	MP2RAGE*
<i>T1w</i>	Voxel size (mm)	0.9x0.9x1	0.9x0.9x1	1x1x1	0.8x0.8x0.8 or 0.6x0.6x0.6
	Flip angle (°)	8	9	12	7 or 5
	TR/TE	8.2/3.7	6.008/1.952	2000/2.52	6000/2.64 or 2.05
	Slice number	170	120	192	176 or 320
	Matrix	256x256	200x200	256x256	240x256 or 256x320
<i>WMn</i>	Voxel size (mm)	0.68x0.68x0.7	0.9x0.9x1.3	1x1x1	*
	Flip angle (°)	8	7	7	
	TR/TE	9.7/4.7	11/5	4000/3.75	
	Slice number	120	210	160	
	Matrix	336x336	200x200	256x256	

Supplementary Table 1: Details of sequence parameters for MRI datasets used. \*Images were synthesized using the inversion recovery equation  $(1 - 2 \exp(-TI/T1))$  with  $TI=670\text{ms}$  and T1w images are generated using the T1 maps produced from MP2RAGE acquisition.

		Linear	2 <sup>nd</sup> order	3 <sup>rd</sup> order	4 <sup>th</sup> order
PHILIPS 3T	SSI	0.53	<b>0.66</b>	<b>0.66</b>	0.64
	MSE	0.06	0.02	0.02	0.02
GE 3T	SSI	0.53	<b>0.71</b>	<b>0.71</b>	0.65
	MSE	0.08	0.02	0.02	0.03
SIEMENS 3T	SSI	0.48	<b>0.63</b>	<b>0.63</b>	0.57
	MSE	0.09	0.04	0.04	0.05
SIEMENS 7T	SSI	<b>0.79</b>	0.72	0.71	0.69
	MSE	0.02	0.04	0.04	0.04

Supplementary Table 2: Mean SSI (structural similarity index) and MSE (mean square error) between the synthesized WMn image and the original WMn image of the same subject by applying a linear (1-x) or averaged 2<sup>nd</sup> order, 3<sup>rd</sup> order or 4<sup>th</sup> order function on its T1w image. This was applied on 10 Philips 3T MPAGE, 9 GE 3T SPGR, 10 Siemens 3T MPAGE and 8 Siemens 7T MP2RAGE cases. The aggregated function is computed by fitting the equation on 10 Philips cases allowing more generalization on other data.



SIEMENS 3T					
	T1w-THOMAS	HIPS-THOMAS	Impr.	CNN	Impr.
<b>THAL</b>	0.91 +/- 0.01	*** 0.93 +/- 0.01	2%	0.91 +/- 0.04	1%
<b>AV</b>	0.67 +/- 0.10	** 0.73 +/- 0.07	7%	0.71 +/- 0.18	6%
<b>VA</b>	0.69 +/- 0.03	** 0.77 +/- 0.05	12%	0.72 +/- 0.09	3%
<b>VLa</b>	0.61 +/- 0.06	0.64 +/- 0.07	5%	0.61 +/- 0.12	0%
<b>VLP</b>	0.76 +/- 0.04	*** 0.86 +/- 0.02	13%	0.81 +/- 0.06	6%
<b>VPL</b>	0.52 +/- 0.15	** 0.76 +/- 0.03	46%	* 0.70 +/- 0.11	33%
<b>Pul</b>	0.85 +/- 0.03	*** 0.88 +/- 0.02	4%	0.86 +/- 0.05	2%
<b>LGN</b>	0.69 +/- 0.08	*** 0.77 +/- 0.05	12%	0.67 +/- 0.17	-2%
<b>MGN</b>	0.76 +/- 0.04	0.78 +/- 0.03	3%	0.70 +/- 0.20	-8%
<b>CM</b>	0.72 +/- 0.05	0.76 +/- 0.03	6%	0.71 +/- 0.14	-2%
<b>MD-Pf</b>	0.83 +/- 0.05	* 0.88 +/- 0.03	6%	0.85 +/- 0.08	3%
<b>Hb</b>	0.66 +/- 0.04	0.66 +/- 0.06	0%	0.60 +/- 0.20	-9%
<b>MTT</b>	0.50 +/- 0.08	* 0.63 +/- 0.05	24%	*** 0.29 +/- 0.11	-42%

Supplementary Table 3: Mean Dice +/- SD of T1w-THOMAS, HIPS-THOMAS, and CNN segmentations along with % improvement for HIPS-THOMAS and CNN compared to T1w-THOMAS for 3T Siemens MPRAGE (n=12) data. HIPS-THOMAS improves Dice by >10% in 4 nuclei and MTT compared to 1 nucleus for CNN. P-values \*<0.00385 (Bonferonni correction for multiple comparisons, 0.05/13) \*\*<0.001 \*\*\*<0.0001.

GE 3T					
	T1w-THOMAS	HIPS-THOMAS	Impr.	CNN	Impr.
<b>THAL</b>	0.91 +/- 0.01	*** 0.93 +/- 0.01	2%	0.92 +/- 0.01	1%
<b>AV</b>	0.68 +/- 0.09	0.69 +/- 0.12	2%	0.70 +/- 0.08	4%
<b>VA</b>	0.70 +/- 0.06	*** 0.81 +/- 0.02	15%	* 0.76 +/- 0.03	12%
<b>VLa</b>	0.64 +/- 0.08	0.66 +/- 0.06	4%	0.62 +/- 0.08	-2%
<b>VLP</b>	0.74 +/- 0.08	*** 0.85 +/- 0.03	16%	*** 0.81 +/- 0.04	12%
<b>VPL</b>	0.56 +/- 0.14	*** 0.78 +/- 0.06	40%	*** 0.72 +/- 0.07	40%
<b>Pul</b>	0.85 +/- 0.03	*** 0.89 +/- 0.02	5%	0.87 +/- 0.02	2%
<b>LGN</b>	0.67 +/- 0.04	*** 0.78 +/- 0.04	17%	*** 0.75 +/- 0.04	13%
<b>MGN</b>	0.78 +/- 0.04	0.82 +/- 0.04	5%	0.69 +/- 0.10	-9%
<b>CM</b>	0.67 +/- 0.09	*** 0.77 +/- 0.06	15%	* 0.75 +/- 0.03	11%
<b>MD-Pf</b>	0.81 +/- 0.05	*** 0.88 +/- 0.02	8%	* 0.86 +/- 0.02	6%
<b>Hb</b>	0.66 +/- 0.07	0.70 +/- 0.04	6%	0.67 +/- 0.12	3%
<b>MTT</b>	0.47 +/- 0.10	*** 0.63 +/- 0.05	32%	** 0.31 +/- 0.08	-34%

Supplementary Table 4: Mean Dice +/- SD of T1w-THOMAS, HIPS-THOMAS, and CNN segmentations along with % improvement for HIPS-THOMAS and CNN compared to T1w-THOMAS for GE 3T SPGR data (n=19). HIPS-THOMAS improves Dice by >15% in 5 nuclei and MTT compared to 1 nucleus for CNN. P-values \*<0.00385 (Bonferonni correction for multiple comparisons, 0.05/13) \*\*<0.001 \*\*\*<0.0001.

	PHILIPS 3T			SIEMENS 7T			SIEMENS 3T
	T1w-THOMAS	HIPS-THOMAS	Impr.	T1w-THOMAS	HIPS-THOMAS	Impr.	HIPS-THOMAS
<b>THAL</b>	0.92 +/- 0.01	*** 0.94 +/- 0.00	2%	0.92 +/- 0.02	** 0.96 +/- 0.02	5%	*** 0.93 +/- 0.01
<b>AV</b>	0.72 +/- 0.06	*** 0.80 +/- 0.03	11%	0.64 +/- 0.16	* 0.84 +/- 0.08	32%	0.73 +/- 0.10
<b>VA</b>	0.71 +/- 0.07	*** 0.82 +/- 0.02	16%	0.63 +/- 0.15	* 0.85 +/- 0.06	35%	0.80 +/- 0.05
<b>VLa</b>	0.58 +/- 0.12	0.64 +/- 0.07	10%	0.48 +/- 0.12	** 0.81 +/- 0.14	69%	* 0.52 +/- 0.15
<b>VLP</b>	0.76 +/- 0.19	0.88 +/- 0.01	16%	0.70 +/- 0.10	** 0.92 +/- 0.04	32%	* 0.87 +/- 0.02
<b>VPL</b>	0.73 +/- 0.07	*** 0.81 +/- 0.02	11%	0.68 +/- 0.10	** 0.90 +/- 0.05	33%	* 0.80 +/- 0.03
<b>Pul</b>	0.87 +/- 0.02	** 0.89 +/- 0.02	2%	0.88 +/- 0.03	* 0.93 +/- 0.03	5%	* 0.88 +/- 0.02
<b>LGN</b>	0.74 +/- 0.04	*** 0.79 +/- 0.03	8%	0.77 +/- 0.05	** 0.88 +/- 0.06	14%	*** 0.76 +/- 0.02
<b>MGN</b>	0.77 +/- 0.04	** 0.80 +/- 0.03	3%	0.79 +/- 0.08	0.86 +/- 0.08	9%	0.75 +/- 0.05
<b>CM</b>	0.70 +/- 0.08	*** 0.79 +/- 0.04	13%	0.51 +/- 0.20	** 0.88 +/- 0.06	73%	* 0.74 +/- 0.08
<b>MD-Pf</b>	0.76 +/- 0.07	*** 0.89 +/- 0.01	17%	0.59 +/- 0.20	* 0.92 +/- 0.03	58%	** 0.86 +/- 0.03
<b>Hb</b>	0.58 +/- 0.08	0.59 +/- 0.04	1%	0.69 +/- 0.12	0.63 +/- 0.15	-10%	0.64 +/- 0.07
<b>MTT</b>	0.48 +/- 0.08	*** 0.66 +/- 0.03	38%	0.54 +/- 0.16	* 0.78 +/- 0.10	46%	** 0.58 +/- 0.07

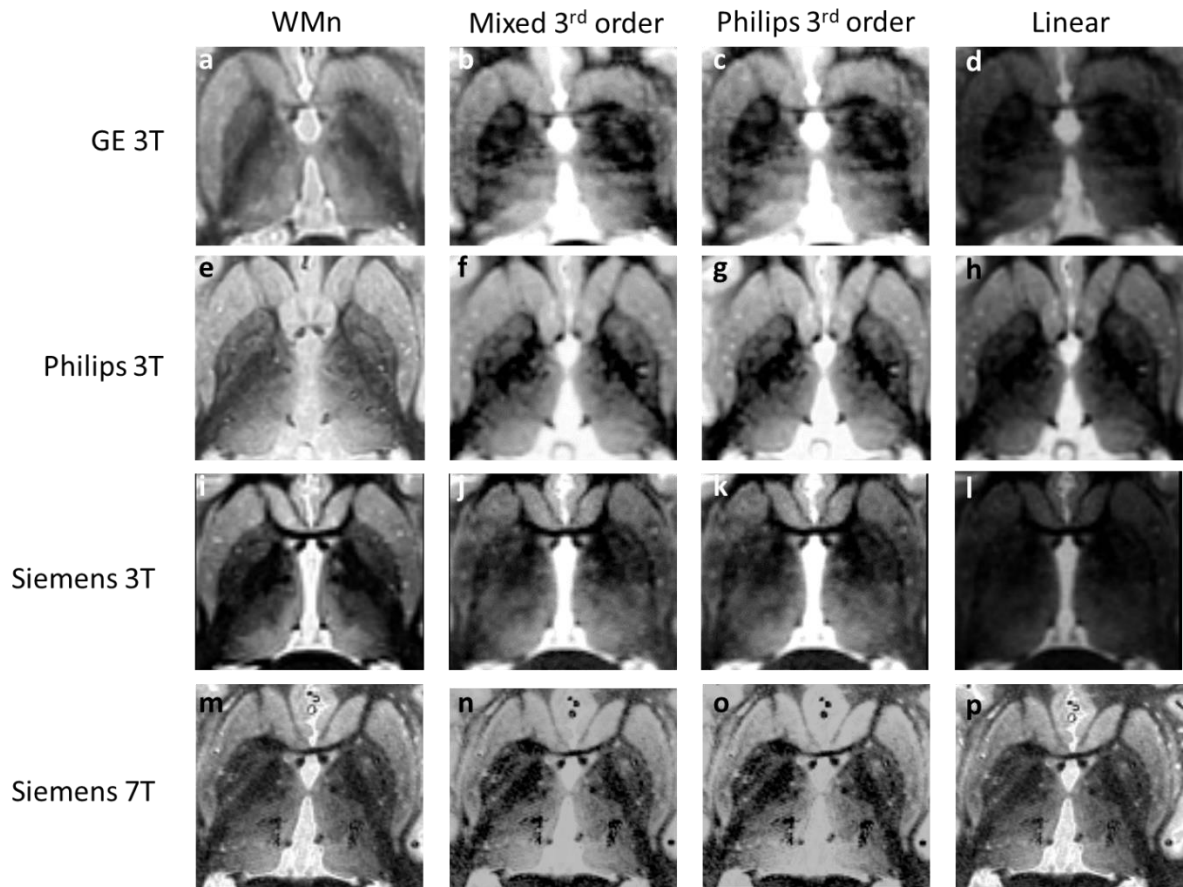
Supplementary Table 5: Mean Dice coefficient +/- SD of T1w-THOMAS and HIPS-THOMAS segmentations along with % improvement for HIPS-THOMAS compared to T1w-THOMAS for Philips 3T MPRAGE (n=18) and Siemens 7T MP2RAGE (n=8) data. HIPS-THOMAS improves Dice by >10% in 6 nuclei and MTT on Philips 3T data and by >30% in 7 nuclei and MTT on Siemens 7T data. The rightmost column shows Dice results for Siemens 3T MPRAGE data obtained on the same 8 subjects scanned on Siemens 7T. P-values \*<0.00385 (Bonferonni correction for multiple comparisons, 0.05/13) \*\*<0.001 \*\*\*<0.0001.

		<i>Thal</i>	<i>AV</i>	<i>VA</i>	<i>Vla</i>	<i>VLp</i>	<i>VPL</i>	<i>Pul</i>	<i>LGN</i>	<i>MGN</i>	<i>CM</i>	<i>MD</i>	<i>Hb</i>	<i>MTT</i>
<i>INTRA</i> ( <i>mm</i> <sup>3</sup> )	T1w	53.8	<b>6.0</b>	<b>8.8</b>	5.6	20.8	<b>12.2</b>	<b>25.4</b>	<b>3.7</b>	<b>3.2</b>	<b>5.2</b>	<b>16.0</b>	<b>2.1</b>	2.8
	HIPS	<b>48.3</b>	7.9	<b>8.8</b>	<b>4.2</b>	<b>19.1</b>	13.4	29.7	6.8	<b>3.1</b>	9.0	24.2	2.6	2.9
	CNN	61.8	9.0	10.6	5.7	24.9	12.5	31.4	5.7	4.0	8.5	17.9	2.5	<b>2.7</b>
<i>INTER</i> ( <i>mm</i> <sup>3</sup> )	T1w	187.2	<b>7.7</b>	18.0	8.2	43.6	26.5	84.1	<b>6.6</b>	6.2	<b>8.9</b>	48.6	<b>3.2</b>	<b>3.3</b>
	HIPS	<b>110.0</b>	13.4	<b>13.4</b>	<b>4.6</b>	<b>37.0</b>	<b>20.7</b>	<b>62.6</b>	11.0	<b>4.3</b>	<b>9.1</b>	<b>30.7</b>	3.7	4.3
	CNN	158.3	28.1	20.8	13.0	58.5	28.7	96.9	13.5	7.8	15.6	51.4	4.1	5.1

Supplementary Table 6: Intra-scanner and inter-scanner variability methods for the Frequently Traveling Human Phantom MRI dataset [5] reported as standard deviation (in units of mm<sup>3</sup>) for T1w-THOMAS, HIPS-THOMAS, and CNN segmentation. For this analysis, 24 different scanners covering 3 manufacturers, 2 field strengths and 4 sites each with 3 repeat scans per site were used, resulting in a dataset comprised of 72 scans.

		Mixed 2 <sup>nd</sup> order	Mixed 3 <sup>rd</sup> order	Philips 2 <sup>nd</sup> order	Philips 3 <sup>rd</sup> order
PHILIPS 3T	SSI	0.66	0.67	0.66	0.66
	MSE	0.02	0.04	0.02	0.02
GE 3T	SSI	0.69	0.70	0.71	0.71
	MSE	0.02	0.02	0.02	0.02
SIEMENS 3T	SSI	0.61	0.61	0.63	0.63
	MSE	0.03	0.03	0.04	0.04
SIEMENS 7T	SSI	0.75	0.77	0.72	0.71
	MSE	0.03	0.03	0.04	0.04

Supplementary Table 7: Mean Structural Similarity Index (SSI) and Mean Square Error (MSE) measures when comparing an axial slice native WMn and synthesized WMn images after applying a quadratic and cubic functions estimated from a mixed dataset comprising of 10 Philips 3T MPRAGE, 9 GE 3T SPGR, 10 Siemens 3T MPRAGE and 8 Siemens 7T MP2RAGE cases. The aggregated function is computed by fitting the equation on mixed data. For comparison, the SSI and MSE from using just 2<sup>nd</sup> and 3<sup>rd</sup> order Philips 3T cases is shown. The two functions perform comparably.



Supplementary Figure 1: An axial slice from native WMn-MPRAGE compared with WMn-like images resulting from the application of the Philips averaged function, mixed dataset function (see Supplementary Table 7 for details), or linear (1-x) function on an example GE 3T SPGR (a-d), Philips 3T MPRAGE (e-h), Siemens 3T MPRAGE (i-l) or Siemens 7T MP2RAGE (m-p) subject. Note the similarity of the transformation when applying the mixed (b, f, j, n) and Philips (c, g, k, o) function on native T1w-images. Note also the optimality of the linear function for Siemens 7T data (p).

## Supplementary References

- [1] Unpublished data from Danet, L., Péran, P., Pariente, J., and Barbeau, E.
- [2] Pfefferbaum, A., Sullivan, E. V., Zahr, N. M., Pohl, K. M., & Saranathan, M. (2023). Multi-atlas thalamic nuclei segmentation on standard T1-weighted MRI with application to normal aging. *Human brain mapping*, 44(2), 612–628.
- [3] Iglehart, C., Monti, M., Cain, J., Tourdias, T., & Saranathan, M. (2020). A systematic comparison of structural-, structural connectivity-, and functional connectivity-based thalamus parcellation techniques. *Brain Structure and Function*, 225, 1631-1642.
- [4] Najdenovska, E. et al. Comparison of MRI-based automated segmentation methods and functional neurosurgery targeting with direct visualization of the Ventro-intermediate thalamic nucleus at 7T. *Scientific reports*, 9(1), 1119 (2019)
- [5] Opfer, R., Krüger, J., Spies, L., Ostwaldt, A. C., Kitzler, H. H., Schippling, S., & Buchert, R. (2023). Automatic segmentation of the thalamus using a massively trained 3D convolutional neural network: higher sensitivity for the detection of reduced thalamus volume by improved inter-scanner stability. *European radiology*, 33(3), 1852–1861.