Multi-Task Learning based 3-Dimensional Striatal Segmentation of MRI – a Multi-modal Objective Assessment

Mario Serrano-Sosa¹, Jared X. Van Snellenberg^{1,2}, Jiayan Meng², Karl Spuhler^{1,#}, Jodi J. Weinstein², Anissa Abi-Dargham², Mark Slifstein², Chuan Huang^{1,2,3,*}

¹Department of Biomedical Engineering, Stony Brook University, Stony Brook, NY

²Department of Psychiatry, Stony Brook Medicine, Stony Brook, NY

³Department of Radiology, Stony Brook Medicine, Stony Brook, NY

#Current Affiliation: NYU Langone, New York, NY

* Please send correspondence to:

Chuan Huang, PhD

HSC-L4-120

Stony Brook Medicine

101 Nicolls Road

Stony Brook, NY 11794

Tel: +1-631-444-6905

Email: chuan.huang@stonybrookmedicine.edu

Highlights:

- Multi-task learning (MTL) model developed for accurate striatal subregion segmentation
- Dice coefficients show high reliability across all five striatal subregions
- PET quantitative analysis shows high agreement between MTL and manual segmentation
- fMRI time-series/whole-brain connectivity show MTL provides reliable quantification

Abstract:

The human striatum is a collection of subcortical nuclei that serves as a critical node in an information processing network. Recent studies have established a clear topographical and functional organization of projections to and from striatal regions that involves detailed and complex subdivisions of the striatum. Manual segmentation of these functional subdivisions is labor-intensive and time-consuming, yet automated methods have also been a challenging computational problem. Recently, deep learning has emerged as a powerful tool for various tasks, including semantic segmentation. Multi-Task Learning (MTL) is a machine learning technique that allows latent representations between related tasks to be shared during model prediction of multiple tasks with better accuracy. We utilized MTL to segment subregions of the striatum consisting of pre-commissural putamen (prePU), pre-commissural caudate (preCA), post-commissural putamen (postPU), post-commissural caudate (postCA), and ventral striatum (VST). Dice similarity coefficients (DSC) demonstrate excellent spatial agreement between manual and MTL-generated segmentations (≥ 0.72 across all striatal subregions). Further quantitative task-based analysis was also conducted. Binding potential values, BP_{ND} , of [^{11}C]raclopride PET, and ROI time-series and whole-brain connectivity using fMRI images were compared between results generated from manual segmentations and MTL-generated segmentations. BP_{ND} values from MTL-generated segmentations were shown to correlate well with manual segmentations with $R_2 \geq 0.91$ in all caudate and putamen subregions, and $R_2=0.69$ in VST. Mean Pearson correlation coefficients of the fMRI data between MTL-generated and manual segmentations were also high in time-series (≥ 0.86) and whole-brain connectivity (≥ 0.89) across all subregions. We conclude that the MTL approach is a fast, robust and reliable method for 3D striatal subregion segmentation with results comparable to manually segmented ROIs.

1. Introduction

The human striatum is a collection of subcortical nuclei that serves as a critical node in an information processing network proceeding from cortex, through the striatum and other basal ganglia structures (e.g. globus pallidus and subthalamic nucleus), to the thalamus and back to cortex. Striatal pathology or dysfunction is thought to play a critical role in neurological and neuropsychiatric diseases such as schizophrenia, Parkinson's, and Huntington's disease¹⁻⁴. Functional and molecular imaging studies of patients with these disorders frequently require identifying the portion of an image volume that contains the striatum and then segmenting the striatum into its anatomical or functional subregions^{5,6}. Automated segmentation of the striatum has been a challenging computational problem, and some researchers continue to rely on manual delineations that are labor-intensive and time-consuming. A plethora of automatic and semi-automatic methods have been proposed to segment brain regions including the striatum. Widely available atlas-based approaches, such as Freesurfer, can robustly identify the whole striatum and its major structures (the caudate and putamen), but have been less reliable in segmenting the striatal subregions, particularly in patient populations⁷.

More recently, deep learning has emerged as a powerful tool for various tasks, including semantic segmentation. Deep learning has the capacity to learn hierarchical features in a data-driven manner through its numerous convolutional neural network (CNN) layers that represent different levels of abstraction. Automated deep learning techniques for semantic segmentation of brain regions from Magnetic Resonance Imaging (MRI) has been recently used to segment subcortical regions, including the striatum and its major structures, the caudate and putamen^{8,9}; however, performance was mainly evaluated using Dice Similarity Score (DSC) with no task-based analysis to determine the accuracy of their segmentations for research use. This latter deep learning method⁹ for striatal segmentation also used two separate networks, a Global CNN to localize the striatum and a Local CNN to further confirm if the chosen voxels were striatal. Although this method provided decent results, more robust deep learning methods exist that do not require multi-network segmentation. Specifically, segmentation using the U-Net architecture has been shown to capture complex feature

context from the contracting path and to enable precise localization from the symmetric expanding path¹⁰. More importantly, no deep learning technique for subregion segmentation has been reported yet.

Anatomical tracing studies in nonhuman primates have established a clear topographical and functional organization of projections to and from striatal regions that involves more detailed and complex subdivisions of the striatum than gross anatomical divisions would suggest¹¹, such that there may be clear benefits to neuroimaging studies if this topography can be captured and utilized in data analysis. Imaging researchers, particularly in positron emission tomography (PET), have translated this topographical organization into operationally defined substructures that can be identified and manually traced on T1-weighted or multi-modal MR images^{12,13}. These approaches require some visual identification and manual outlining of substructures by analysts with expert knowledge and are time consuming, requiring a day or more to complete a single striatum. As a result, functional Magnetic Resonance Imaging (fMRI) studies very rarely, if ever, use manual segmentation, with most studies employing a set of several “seed” regions of interest that are implemented as small spheres drawn in a standardized coordinate space in the striatum¹⁴⁻¹⁸. However, this approach fails to take any account individual differences in striatal topography and may fail to optimally capture meaningful subdivisions in striatal structures.

Multi-Task Learning (MTL) is a technique used in deep learning that allows representations between related tasks to be shared that can generalize the model to predict the original task, and many more, with better accuracy¹⁹. Also called joint learning, learning to learn, and learning with auxiliary tasks, this method has been known to optimize multiple tasks and reduce overfitting^{19,20}. In this study we sought to utilize MTL to segment subregions of the striatum consisting of pre-commissural putamen (prePU), pre-commissural caudate (preCA), post-commissural putamen (postPU), post-commissural caudate (postCA), and ventral striatum (VST) using a convolutional neural network (CNN). More specifically, we employ a 3D U-Net architecture with single input and multiple outputs to segment the striatal regions of interest (ROIs) and background. Here, we assess these striatal segmentations with DSC with respect to the gold standard manually drawn ROIs. More importantly, we use

task-based analysis applied to the MTL segmentations to re-analyze previously published PET binding potential data, as well as resting state functional connectivity (RSFC) fMRI data, obtained from manually drawn ROIs²¹ and compare these results to those obtained with our new methodology.

2. Methods

2.1 Datasets

Table 1. Details of the T1-weight MRI data used in this study.

	Training set 1	Training set 2	testing set
Sample size	46*	22*	19
Age (SD)	32.5(10.2)	31.0 (12.1)	33.2(9.7)
Sex (M/F)	21/25	15/7	13/6
Scanner	Philips Intera 1.5T	GE MR 750 3T	GE MR 750 3T
Original publication	Horga et al., 2016; Van Snellenberg et al., 2016 ^{22,23}	Ongoing, unpublished	Weinstein et al., 2018 ²¹
Sequence	SPGR SENSE	3D BRAVO	3D BRAVO
TR (ms)	25	7.82	7.82
TE (ms)	5.67	3.09	3.09
Parallel imaging factor	1.5	2	2
Voxel size (mm)	1	0.8	0.8
Flip angle	30	12	12

* 1 and 3 subjects out of training set 1 and 2, respectively, were used as validation

2.1.1 Training/Validation Data

Sixty-eight data sets from both previously published^{22,23} and ongoing studies of patients with schizophrenia and matched controls were employed. Striatal and extrastriatal ROIs were manually drawn on individual T1-weighted scans using a previously validated method^{24,25}, which is a time-consuming procedure. The 3D T1-weighted image was used as input for the MTL network and six output tasks were

simultaneously trained; where five of the six were ROI masks and the last was the masked background. 68 total datasets were split into 64 training and 4 validation sets (See Table 1).

2.1.2 PET Test Data set

We used data from a previously published study examining dopamine D2 receptor internalization in 10 patients with schizophrenia and 9 matched healthy controls²¹, which were independent from the training datasets. PET data were acquired on a Siemens mCT scanner. T1-weighted anatomical images for ROI delineation were acquired for each subject on a GE750 3T scanner. The PET data were corrected for motion and coregistered to the subjects' MRI data using the normalized maximization of mutual information algorithm in SPM software (Wellcome Trust, Cambridge, UK). ROIs were manually delineated on the MR images as in Mawlawi et al.¹² and applied to the coregistered PET. 60 min of dynamic emission data were acquired, following a bolus injection of 349 ± 109 MBq of [^{11}C]raclopride. Data were reconstructed by filtered back projection (FBP) with computed tomography (CT) used for attenuation correction.

2.1.3 fMRI Test Dataset

A subset of 16 participants (7 patients and 9 controls) from the PET test dataset described above also underwent 30 minutes of RSFC scanning and made up the fMRI test dataset. The fMRI scanning was conducted on the same GE MR 750 3T scanner. Multiband blood oxygen level dependent (BOLD) MR sequences were acquired using a multiband acceleration factor of 6, no in-plane acceleration or parallel imaging, with 66 slices, 192 mm field-of-view, 2 mm isotropic voxel size, 60° flip angle, 850 ms TR, and 25 ms TE. Each participant was scanned in four runs with alternating AP and PA phase encode directions, with each run consisting of 532 volumes. B_0 field maps were acquired along with gradient echo images in both phase encode directions in the same geometry as the BOLD sequences (B_0 maps were preferentially used for distortion correction, however in cases where distortion correction failed gradient echo images were used instead).

2.2 Network architecture

As shown in Figure 1, this MTL network consisted of a 3D U-Net architecture with an encoding path (contracting, left side) and decoding path (expanding, right side). Images were initially cropped to ensure that the striatum was used as input. The encoding path conforms to the typical architecture of a convolutional neural network (CNN), consisting of the repeated application of three $3 \times 3 \times 3$ convolution layers, each followed by an exponential linear activation (elu). Each block ended with a $2 \times 2 \times 2$ maxpooling layer with stride 2 for downsampling. In addition, at each downsampling step, the number of feature channels was doubled in the encoder path; this increment was initialized at 16 feature maps in the first convolutional block and increased to a maximum of 128. The decoding path consisted of similar convolutional blocks with three $3 \times 3 \times 3$ convolutional layers ending with elu but contain $3 \times 3 \times 3$ convolutional transpose with stride of 2 for upsampling instead of the previously used maxpooling. In this decoding path, the feature channel was halved and skip connection with the corresponding linked feature map from the encoding path was utilized. The final step was a $1 \times 1 \times 1$ convolution that maps the output classification.

Figure 1: Schematic of the U-Net architecture used for this Multi-Task Learning based segmentation.

2.3 Model Training

This model was trained to minimize the sparse softmax cross entropy between output and manually drawn ROIs; where the function first uses softmax to transform the outputs into a probability distribution and computes the loss function:

$$L(w) = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)],$$

where w refers to the model parameters, e.g. weights of the neural network, y_i is the true label, and $\hat{y}_i$ is the predicted label.

The training set and a validation set, together consisting of 68 datasets, was used to optimize the hyperparameters of the model. The model was trained using the Adam optimizer and non-decaying learning rate of 1×10^{-5} ; network parameters were initialized using the Glorot method²⁶. The model was trained on a computer with an i7-8700K 6-core processor, 32 GB memory, and GTX1080 Ti graphics cards running Ubuntu 18.04, Python 2.7.15, and Tensorflow 1.11.0.

2.4 Segmentation Evaluation

DSC was utilized to initially evaluate the performance of the MTL network output:

$$DSC = \frac{2(V_A \cap V_B)}{V_A + V_B},$$

where A is the automatic segmentation from the MTL network and B is the hand drawn ROI.

2.5 Task based Assessment

After generating striatal subregion segmentations from MTL network on the test data sets, task-based analyses were conducted on PET and fMRI imaging to determine the reliability of MTL-generated compared to manual segmentations.

2.5.1 Quantitative PET Analysis

In accordance with the original publication²¹ of the PET test set, the primary outcome measure was BP_{ND}, the binding potential with respect to the non-displaceable compartment²⁷. BP_{ND} was assessed in each ROI using the simplified reference tissue model²⁸ (SRTM) with cerebellum as a reference tissue. Here, we compared baseline BP_{ND} values in all 19 subjects from Weinstein et al.²¹ to BP_{ND} as assessed using MTL-generated ROIs applied to the subjects' T1-weighted images and the original coregistered PET data. The kinetic modeling procedure was the same as in the original publication. Linear Regression was performed with BP_{ND} obtained from manual and

MTL-generated ROIs. R^2 and percent differences are reported for manual and MTL-generated ROIs. Intraclass correlation coefficients (ICC(3,2): 2-way mixed model, two-rater, consistency) were also calculated for all subregions with BP_{ND} from MTL-generated segmentations relative to BP_{ND} from manual segmentations.

2.5.2 fMRI Connectivity Analysis

RSFC data was initially preprocessed using the Human Connectome Project (HCP) minimal preprocessing pipeline²⁹, and the first 7 volumes of each run were discarded to reach steady-state. Data from one patient participant could not be appropriately distortion-corrected, leaving substantial spatial distortions in the data relative to other participants, and so data from this participant was excluded. Both sets of CNNs were coregistered to RSFC data using SPM 12 spatial coregistration of individual subject T1w images used for manual tracings to the individual-subject space T1w images output by the HCP pipeline. Raw timeseries were then extracted from both hand-drawn and MTL-generated ROIs following preprocessing, and the correlation between the hand-drawn and CNN timeseries data was calculated for each fMRI run. These correlations were then transformed using Fisher's r-to-Z transformation, averaged across runs, and converted back to r values for presentation.

Next, whole-brain data in MNI space was spatially smoothed using a 4 mm full-width-half-maximum (FWHM) kernel in SPM 12 and underwent standard RSFC processing along with each ROI timeseries in a custom Matlab pipeline: 1) linear detrending and de-meaning of each timeseries, 2) regression of average gray matter, white matter, and CSF signals, 3) regression of motion parameters, their squares, their first derivatives, and their squared first derivatives ("Friston-24" motion parameter regression³⁰), 4) identification of "bad" volumes for censoring using recently-developed methods for multiband RSFC data that we have described in detail elsewhere³¹, 5) linear interpolation of signal over bad volumes, 6) bandpass filtering from 0.09 – 0.008 Hz, 7) discarding of the first and last 30 volumes of each timeseries to eliminate distortion due to digital filtering, and finally 8) volume censoring. Timeseries correlations between each voxel were calculated for each ROI and averaged across runs to obtain

whole-brain RSFC correlations. Finally, the correlation across voxels in RSFC was calculated between hand-drawn and MTL-generated ROIs.

3. Results

The CNN was successfully trained to segment all ROIs. Relative to manually based segmentations that take 1.5 days per subject, our MTL-generated segmentation takes 30 seconds to generate five 3D masks. Figure 2 shows an example of both manual and MTL-generated segmentations; visually, MTL-generated segmentations are similar to manually segmented ROIs. As shown in Figure 2, MTL-generated segmentation seems to smooth VST where manual segmentation may have sharp edges.

Figure 2: Transverse (top), coronal (middle) and sagittal (bottom) views of T1-weighted MR Image (left), overlaid with ROIs by manual segmentation (middle) and MTL-generated segmentation (right). Transverse slices of manual segmentation show left postPU to have irregular shape compared to a smoother left postPU in MTL-generated segmentation. Coronal and sagittal slices of the MRI and manual segmentation also show VST to have irregular shape compared to a smoothed and more circular VST in MTL-generated.

We first tested our MTL-generated segmentations by comparing to manually drawn ROIs using DSC on a completely independent testing set. Table 1 shows mean DSC across the independent testing set with 0.8123 ± 0.0795 , 0.8716 ± 0.0364 , 0.8234 ± 0.0348 , 0.7506 ± 0.0485 , 0.7221 ± 0.0643 , for prePU, preCA, postPU, postCA and VST, respectively.

PET quantitative analysis showed R^2 between BP_{ND} values using manual and MTL-generated ROIs was ≥ 0.91 in all subdivisions of the caudate and putamen, and 0.69 in VST (Figure 3). Table 2 shows regression coefficients for this analysis. The slopes were all close to 1, and the intercepts were relatively small. Figure 4 shows the percent differences of BP_{ND} between the manual and MTL-generated ROIs. Although CNN segmentation of postCA had regression coefficient of 1 and a high R^2 of 0.91, Figure 4 shows that it had a downward bias. ICC for all subregions was also ≥ 0.835 as shown in table 2.

fMRI results were similar, as shown in Table 3 and Figure 5, which show mean Pearson correlation coefficients between manual and MTL-generated segmentations for fMRI timeseries and whole-brain connectivity analysis. fMRI timeseries data prior to RSFC processing were correlated at 0.95 or above for preCA and both putamen subregions, with somewhat lower values of 0.87 and 0.85 for postCA and VST, respectively (RSFC processed timeseries data are not shown, but correlations differed by no more than 0.01 for all subregions). Whole-brain connectivity values were similarly highly correlated, with r values of 0.96 or higher for preCA and both putamen subregions, and of 0.91 and 0.88 for postCA and VST, respectively.

Table 1: Mean Dice Similarity Coefficient (DSC) with standard deviations for all sub-region ROIs

	prePU	preCA	postPU	postCA	VST
DSC	$0.8123 \pm$	$0.8716 \pm$	$0.8234 \pm$	$0.7506 \pm$	$0.7221 \pm$

	0.0795	0.0364	0.0348	0.0485	0.0643
--	--------	--------	--------	--------	--------

Table 2: Regression coefficients from correlative analysis between BP_{ND} calculated from manually drawn ROIs and MTL-generated ROIs. R_2 from regression analysis, percent difference with standard deviations, and ICC comparing BP_{ND} from CNN generated and manual segmentations are also shown.

	prePU	preCA	postPU	postCA	VST
slope	1.04	1.07	1.11	1.00	0.93
intercept	-0.08	-0.16	-0.44	-0.17	0.18
R₂	0.93	0.93	0.96	0.91	0.69
% Diff.	1.53±2.45	0.54±3.12	-2.41±2.82	-9.83±6.22	0.01±4.75
ICC*	0.974	0.979	0.975	0.917	0.910

* ICC(3,2): 2-way mixed model, two-rater, consistency

Figure 3: Scatter plots between BP_{ND} of the 5 subregions calculated using manual ROIs and MTL-generated ROIs across 19 independent test subjects. Regression line for all subregions is overlaid scatter plots, with regression coefficients found in Table 2.

Figure 4: Mean Percent Differences (standard deviation) between BP_{ND} of the 5 subregions calculated using manual ROIs and MTL-generated ROI across 19 independent test subjects.

Table 3: Mean Pearson correlation coefficients (Pearson's *r*) with 95% confidence intervals between manual ROIs and MTL-generated ROIs for the all sub-regions in fMRI timeseries and whole-brain connectivity analysis

	prePU	preCA	postPU	postCA	VST
Timeseries	0.9693 [0.9664,0.9720]	0.9761 [0.9739,0.9781]	0.9503 [0.9472,0.9533]	0.8888 [0.8791,0.8978]	0.8665 [0.8561,0.8762]
Whole-brain connectivity	0.9782 [0.9759,0.9802]	0.9833 [0.9815,0.9850]	0.9671 [0.9636,0.9703]	0.9168 [0.9093,0.9236]	0.8915 [0.8844,0.8981]

Figure 5: Mean Pearson correlation coefficients (Pearson's r) with error bars depicting 95% confidence intervals between manual ROIs and MTL-generated ROIs for the all sub-regions in (left) fMRI timeseries and (right) whole-brain connectivity analysis.

4. Discussion

In this study, we developed a deep MTL framework for striatal subregion segmentation and assessed the network outputs with PET and fMRI-derived quantitative outcome measures. We show here that this method provides extremely fast and reliable segmentations for the purpose of PET quantification, fMRI timeseries extraction, and whole-brain RSFC correlations, of the subregions of the striatum that avoid the labor-intensive task of hand drawing ROIs. This network architecture was designed to optimize the segmentation of multiple striatal subregions trained and validated on a total of 68 3D T1-weighted MR images and their manual segmentations. The network was further evaluated on 19 test datasets for PET and 15 test datasets for fMRI.

Visually, the MTL-generated ROIs look extremely similar to manual segmentations, despite smoothing of edges in the CNN output. DSC scores obtained from the comparison of manual and MTL-generated segmentations were > 0.8 in 3 of the 5 ROIs and > 0.7 across all subregions, with the lowest DSC score associated with the VST. The lower VST score is likely due to its small size and irregular shape.

Nevertheless, comparative PET quantitative analysis between manual and MTL-generated segmentations showed that, although slightly smoothed, our MTL-generated segmentations agree well with the manual segmentation results. Likewise, we observe a downward bias in the postCA, a region that is long, narrow, and surrounded by background, which may make it difficult for the network to discern and segment this region. Nevertheless, R_2 of BP_{ND} between methods was 0.91, and whole-brain RSFC of this region was correlated at 0.91 between manual and CNN ROIs, indicating that PET quantification and RSFC analysis using our network can explain most of the variance in the manually delineated data.

Growing interest in segmenting the striatum into its functional subdivisions led to a previous investigation comparing manual and automatic striatal subregion segmentation for interrater, and intrarater, reliability on measuring BP_{ND} from [¹¹C]raclopride PET³². This study divided the striatum into 3 subdivisions: limbic striatum, consisting of VST; associative striatum, consisting of preCA and prePU; and sensorimotor striatum, consisting of postCA and postPU. This study not only showed large differences between two manual raters, but also showed conventional automated methods such as FreeSurfer to have high test-retest variability (TRV) of up to -15.6% for VST BP_{ND} alone. Interrater reliability between BP_{ND} measured from FreeSurfer limbic striatum segmentations and manual segmentation showed Pearson's r of 0.81. Across associative and sensorimotor striatum, automated methods perform rather well with Pearson's $r > 0.88$ and 0.95³², respectively. We used post-hoc analysis to group our five subregions into limbic, associative and sensorimotor functional groups as mentioned in this publication. Our results showed Pearson's r to be 0.832 in limbic striatum, 0.96 in both substructures of the associative striatum and 0.95 and 0.98 in the substructures of the sensorimotor striatum. This suggests that our MTL-generated segmentations can be used to generate reliable measurements compared to manually based segmentations.

Fast and reliably obtained striatal subregion segmentations are extremely beneficial in research settings, facilitating accurate and reproducible quantification of neuroreceptors, indices of neurotransmission capacity and striatal-cortical connectivity^{1,24,25}; all of which are of strong interest in studies of patients with

schizophrenia, as well as many other neuropsychiatric and neurological disorders. Overall, fast computational speed for striatal segmentations increases feasibility of large brain imaging cohorts for large-scale structural and functional brain studies. The accuracy that our CNN-based segmentation has achieved allows this method to be a very advantageous approach for imaging studies.

An issue that all deep learning models face is the number of training/validation samples used to optimize the network. Our network used 68 subjects to train/validate the network, which is larger than the 13 datasets for deep learning striatal segmentation used in prior work⁹. Moreover, our network was able to generate segmentations for subregions of the striatum, which has not been attempted previously, to our knowledge. Using an MTL approach allows our network to share similar parameters across each ROI from the same T1-weighted input to optimize segmentation, thereby creating a more robust network for segmenting individual ROIs, despite the number of training samples used. Nevertheless, our approach could always benefit from more training datasets to improve performance and avoid the smoothing that the network has shown for the ROIs, especially VST. Other non-deep learning methods did not require manual segmentations and were able to segment regions such as the putamen and caudate³³, using multimodal imaging such as T1-, T2-, and Diffusion-weighted imaging (DWI) from the Human Connectome Project (HCP) dataset. This DWI acquisition from HCP had a total acquisition time of one hour, which is not standard across all MRI protocols. Our network is much more generalizable since it utilizes only T1-weighted images to segment several subregions of the striatum. Nonetheless, the network used here can be easily modified to adapt to include multiple contrasts as input channels. This MTL network is designed to optimize such adaptations by utilizing hard parameter sharing between inputs to segment multiple ROIs at once. This network can also be made to segment other structures of the brain due to the flexibility inherent in our MTL network, especially since it is already trained to segment regular and irregular shapes such as the caudate and ventral striatum. Modifications and re-training using different contrasts and datasets would be required but could easily be implemented for this MTL network.

In this paper, we have developed a fast, robust and reliable method for 3D striatal subregion segmentation using an MTL approach. This deep learning

performance was comparable to manually segmented ROIs but exponentially faster. Moreover, our MTL-generated segmentations were reliable in PET and fMRI quantification. Due to its speed and accuracy in segmenting striatal subregions, this MTL network has broad applicability in neuroimaging.

5. Acknowledgement

The authors would like to thank Xiaoyan Xu PhD for performing the manual segmentation on some of the data sets. Previously published data used in Training Set 1 was funded by P50 MH 086404 (AAD). Unpublished data used in Training Set 2 was funded by K01 MH 107763 (JXVS). Previously published data used in Testing Set 1 was funded by R21 MH 099509 (AAD). JJW was supported by K23 MH 115291, JXVS was supported by K01 MH 107763, and CH was supported by NARSAD 24971.

6. Declaration of Competing Interest

The authors declare they have no conflicts of interest to disclose.

References

1. Horga G, Cassidy CM, Xu X, et al. Dopamine-Related Disruption of Functional Topography of Striatal Connections in Unmedicated Patients With Schizophrenia. *JAMA Psychiatry*. 2016;73(8):862-870.
2. Laruelle M, Abi-Dargham A, Gil R, Kegeles L, Innis R. Increased dopamine transmission in schizophrenia: relationship to illness phases. *Biological psychiatry*. 1999;46(1):56-72.
3. Howes OD, Kambeitz J, Kim E, et al. The nature of dopamine dysfunction in schizophrenia and what this means for treatment: meta-analysis of imaging studies. *Archives of general psychiatry*. 2012;69(8):776-786.
4. Teichmann M, Gaura V, Démonet J-F, et al. Language processing within the striatum: evidence from a PET correlation study in Huntington's disease. *Brain : a journal of neurology*. 2008;131(4):1046-1056.
5. Martinez D, Slifstein M, Broft A, et al. Imaging human mesolimbic dopamine transmission with positron emission tomography. Part II: amphetamine-induced dopamine release in the functional subdivisions of the striatum. *Journal of cerebral blood flow and metabolism : official journal of the International Society of Cerebral Blood Flow and Metabolism*. 2003;23(3):285-300.
6. Mawlawi O, Martinez D, Slifstein M, et al. Imaging human mesolimbic dopamine transmission with PET: I. Accuracy and precision of D2 parameter measurements in the ventral striatum. *Journal of Cerebral Blood Flow and Metabolism*. 2001;21(9):1034-1057.
7. Makowski C, Beland S, Kostopoulos P, et al. Evaluating accuracy of striatal, pallidal, and thalamic segmentation methods: Comparing automated approaches to manual delineation. *Neuroimage*. 2018;170:182-198.
8. Dolz J, Desrosiers C, Ben Ayed I. 3D fully convolutional networks for subcortical segmentation in MRI: A large-scale study. *NeuroImage*. 2018;170:456-470.
9. Choi H, Jin KH. Fast and robust segmentation of the striatum using deep convolutional neural networks. *Journal of Neuroscience Methods*. 2016;274:146-153.
10. Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. Paper presented at: International Conference on Medical Image Computing and Computer-Assisted Intervention2015.
11. Haber SN, Fudge JL, McFarland NR. Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *J Neurosci*. 2000;20(6):2369-2382.
12. Mawlawi O, Martinez D, Slifstein M, et al. Imaging human mesolimbic dopamine transmission with positron emission tomography: I. Accuracy and precision of D-2 receptor parameter measurements in ventral striatum. *J Cerebr Blood F Met*. 2001;21(9):1034-1057.
13. Tziortzi AC, Haber SN, Searle GE, et al. Connectivity-based functional analysis of dopamine release in the striatum using diffusion-weighted MRI and positron emission tomography. *Cerebral cortex*. 2014;24(5):1165-1177.
14. Di Martino A, Kelly C, Grzadzinski R, et al. Aberrant striatal functional connectivity in children with autism. *Biological psychiatry*. 2011;69(9):847-856.

15. Di Martino A, Scheres A, Margulies DS, et al. Functional connectivity of human striatum: a resting state fMRI study. *Cerebral cortex*. 2008;18(12):2735-2747.
16. Sarpal DK, Argyelan M, Robinson DG, et al. Baseline Striatal Functional Connectivity as a Predictor of Response to Antipsychotic Drug Treatment. *The American journal of psychiatry*. 2016;173(1):69-77.
17. Sarpal DK, Robinson DG, Lencz T, et al. Antipsychotic treatment and functional connectivity of the striatum in first-episode schizophrenia. *JAMA psychiatry*. 2015;72(1):5-13.
18. Posner J, Marsh R, Maia TV, Peterson BS, Gruber A, Simpson HB. Reduced functional connectivity within the limbic cortico-striato-thalamo-cortical loop in unmedicated adults with obsessive-compulsive disorder. *Human brain mapping*. 2014;35(6):2852-2860.
19. Ruder S. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*. 2017.
20. Baxter J. A Bayesian/information theoretic model of learning to learn via multiple task sampling. *Machine learning*. 1997;28(1):7-39.
21. Weinstein JJ, van de Giessen E, Rosengard RJ, et al. PET imaging of dopamine-D2 receptor internalization in schizophrenia. *Molecular psychiatry*. 2018;23(6):1506-1511.
22. Horga G, Cassidy CM, Xu X, et al. Dopamine-Related Disruption of Functional Topography of Striatal Connections in Unmedicated Patients With Schizophrenia. *JAMA psychiatry*. 2016;73(8):862-870.
23. Van Snellenberg JX, Girgis RR, Horga G, et al. Mechanisms of Working Memory Impairment in Schizophrenia. *Biological psychiatry*. 2016;80(8):617-626.
24. Abi-Dargham A, Martinez D, Mawlawi O, et al. Measurement of striatal and extrastriatal dopamine D1 receptor binding potential with [11C] NNC 112 in humans: validation and reproducibility. *Journal of Cerebral Blood Flow & Metabolism*. 2000;20(2):225-243.
25. Abi-Dargham A, Xu X, Thompson JL, et al. Increased prefrontal cortical D1 receptors in drug naive patients with schizophrenia: a PET study with [11C] NNC112. *Journal of psychopharmacology*. 2012;26(6):794-805.
26. Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. Paper presented at: Proceedings of the thirteenth international conference on artificial intelligence and statistics 2010.
27. Innis RB, Cunningham VJ, Delforge J, et al. Consensus nomenclature for in vivo imaging of reversibly binding radioligands. *J Cereb Blood Flow Metab*. 2007;27(9):1533-1539.
28. Lammertsma AA, Hume SP. Simplified reference tissue model for PET receptor studies. *NeuroImage*. 1996;4(3 Pt 1):153-158.
29. Glasser MF, Sotiropoulos SN, Wilson JA, et al. The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*. 2013;80:105-124.
30. Friston KJ, Williams S, Howard R, Frackowiak RS, Turner R. Movement-related effects in fMRI time-series. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*. 1996;35(3):346-355.
31. Williams JC, Van Snellenberg JX. Motion denoising of multiband resting state functional connectivity MRI data: An improved volume censoring method. *BioRxiv*. 2019:49.

32. Johansson J, Alakurtti K, Joutsa J, Tohka J, Ruotsalainen U, Rinne JO. Comparison of manual and automatic techniques for substriatal segmentation in 11C-raclopride high-resolution PET studies. *Nuclear medicine communications*. 2016;37(10):1074-1087.
33. Visser E, Keuken MC, Douaud G, et al. Automatic segmentation of the striatum and globus pallidus using MIST: Multimodal Image Segmentation Tool. *NeuroImage*. 2016;125:479-497.