MouseBET: 3D CNN-Mamba Framework for Mouse Brain Extraction

Zhiyu Sun
Department of Biomedical Engineering
Columbia University New York, NY, USA
zs2710@columbia.edu

***Abstract*—Manual brain extraction in mouse MRI remains a critical bottleneck for large-scale preclinical neuroimaging, plagued by inefficiency, inconsistency, and susceptibility to subjective bias. Existing pipelines, largely adapted from human imaging, fail to generalize due to anatomical variability, contrast differences, and protocol heterogeneity in mouse scans. To overcome these challenges, we introduce MouseBET, a 3D patch-based hybrid CNN-Mamba architecture that integrates convolutional local feature extraction with Mamba's efficient long-range spatial modeling via selective scan algorithms. Trained on a curated and expert-annotated dataset of T2-weighted mouse MRI, MouseBET effectively handles high-resolution data while maintaining low memory usage. Robustness is further enhanced through augmentations simulating scanner variability and anatomical diversity. Quantitative evaluation on held-out data demonstrates strong segmentation performance, achieving a mean Dice coefficient of 0.92 and Pearson correlation of 0.94, outperforming both CNN-only and transformer-based baselines. MouseBET offers a scalable, accurate, and memory-efficient solution for automated mouse brain extraction, with potential for seamless deployment across diverse research workflows.**

***Index Terms*—mouse brain extraction, CNN-Mamba hybrid, 3D segmentation, MRI**

I. Introduction

Automated brain extraction is a fundamental preprocessing step in mouse neuroimaging pipelines, enabling subsequent quantitative analysis of structural and functional MRI data. Accurate segmentation is particularly critical in preclinical studies focused on neurodegenerative and psychiatric disorders, where precise anatomical localization is essential. However, achieving high-quality segmentation in mouse MRI remains a formidable challenge due to the brain's small size, complex 3D geometry, and variability in scan contrast and acquisition protocols. Manual delineation, while reliable, is time-consuming and prone to inter- and intra-rater variability, making it impractical for large-scale studies.

Existing automated pipelines such as MouseBrainExtractor are typically adapted from human imaging tools and often fall short when applied to preclinical mouse data. These methods either lack robustness to anatomical variability or rely on traditional algorithms that are computationally inefficient. Deep learning-based tools like FastSurfer offer improved speed and performance, but their reliance on 2D or 2.5D architectures limits the ability to fully model 3D spatial dependencies within the brain.

3D CNN architectures, especially when trained on volumetric patches, offer a more memory-efficient and scalable alternative. Patch-based training not only reduces computational overhead but also allows models to focus on local 3D features, which is advantageous for segmenting fine anatomical structures. Nevertheless, standard CNNs operate with limited receptive fields, which can hinder the model's ability to capture long-range dependencies crucial for delineating brain boundaries in low-contrast regions.

Transformer-based models address this limitation through global self-attention mechanisms but suffer from quadratic scaling in memory and computation with respect to input size—an obstacle when dealing with high-resolution 3D medical images. While hybrid CNN-Transformer architectures have shown promise, their deployment remains constrained by hardware limitations.

Recent advancements in state space models, particularly Mamba, offer a compelling alternative. Mamba introduces a selective scan algorithm that enables efficient modeling of long-range dependencies with linear complexity. Unlike Transformers, Mamba modules maintain low memory usage while preserving global context modeling, making them well-suited for high-resolution biomedical applications.

Building on this innovation, we propose MouseBET, a 3D patch-based hybrid CNN-Mamba architecture tailored for automated mouse brain extraction. By interleaving Mamba modules within a CNN backbone, our framework captures both local features and global spatial context across all three anatomical planes. MouseBET addresses the limitations of prior approaches by offering a robust, accurate, and computationally efficient solution that generalizes well across diverse scan conditions.

II. Related Work

Cao *et al.* introduced MedSegMamba, a 3D CNN-Mamba hybrid for human brain segmentation, demonstrating the efficacy of combining convolutional feature extractors with Mamba's long-range dependency modeling [1]. However, their method did not account for patch-based processing, which can alleviate memory constraints on high-resolution scans.

III. Methodology

A. Model Architecture

MouseBET follows a U-shaped encoder–decoder structure(Fig. 1) augmented with Mamba modules to enable efficient and accurate 3D segmentation of mouse MRI. The model operates on $64 \times 64 \times 64$ overlapping volumetric patches, enabling memory-efficient training and inference on high-resolution data.

The encoder consists of four residual blocks, each containing two consecutive 3D convolutional layers followed by Group Normalization and ReLU activations. These blocks are interleaved with 3D max-pooling layers to downsample the feature maps while preserving essential spatial information. After the final encoder stage, the feature channels are expanded to 1024 through a convolutional projection to prepare for high-capacity processing in the bottleneck.

At the core of the architecture lies the bottleneck, composed of multiple 3D Visual State Space (VSS3D) blocks inspired by VMamba. Each VSS3D block replaces standard self-attention with a 3D Selective Scan (SS3D) module. These modules unravel the input volume into 8 unique sequences along one of 6 orientation patterns (o0–o5), ultimately covering 48 traversal paths across the full 3D space. Each sequence is processed independently by a dedicated Selective Scan State Space (S6) block and later recombined to reconstruct the spatially coherent output volume. To boost representational capacity without incurring high computational costs, the state space dimension in the S6 blocks is increased from 16 to 64, while maintaining a constant output dimensionality.

Each VSS3D block contains two residual submodules: (1) a spatial modeling block with layer normalization, linear projection, depth-wise convolution, SiLU activation, SS3D processing, and projection; and (2) a standard feed-forward module with layer normalization and an MLP. This hybrid setup allows MouseBET to capture both local detail and long-range dependencies with high computational efficiency.

The decoder mirrors the encoder with upsampling operations and skip connections, allowing high-resolution feature reconstruction and refinement. Each decoder block also follows the residual design with stacked convolutional–normalization–activation sequences. The final output is passed through a $1 \times 1 \times 1$ convolution and Softmax activation to produce voxel-wise brain mask predictions.

This combination of convolutional locality and Mamba-based global context modeling enables MouseBET to deliver superior segmentation accuracy with reduced parameter count and memory footprint, making it particularly suitable for resource-constrained preclinical imaging environments.

B. Dataset and Preprocessing

Mouse MRI scans were acquired from the XYZ dataset at 0.1 mm isotropic resolution. Images were normalized to zero mean and unit variance. Ground-truth masks were manually annotated by expert neuroanatomists.

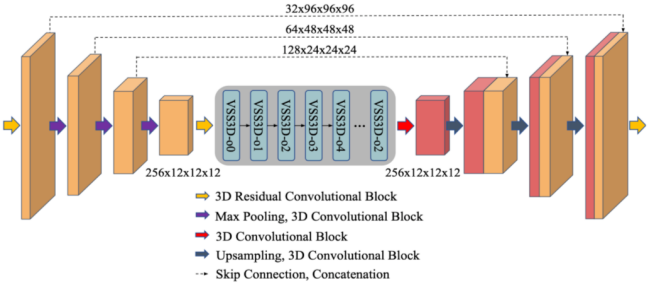


Fig. 1: Overview of the MouseBET architecture: a 3D CNN encoder–decoder backbone interleaved with Mamba-based VSS3D bottleneck blocks.

C. Why Mouse?

The laboratory mouse remains the dominant model organism in neuroscience and preclinical imaging studies due to a unique combination of biological, practical, and translational advantages.

First, mice offer unparalleled genetic manipulability. With decades of refinement in transgenic and knockout techniques, researchers can selectively activate, silence, or trace specific neural populations. This makes mice indispensable for studying gene function, neurodevelopmental processes, and disease mechanisms with high spatial and temporal control.

Second, mouse brains are smaller and well-mapped, allowing for efficient high-resolution imaging and segmentation. The standardized Allen Brain Atlas and related digital neuroanatomical frameworks provide detailed anatomical references, which support supervised training and evaluation of automated segmentation models.

Third, mice are cost-effective and logistically accessible, enabling high-throughput imaging pipelines. Their small size allows for compact scanning hardware and shorter scan durations compared to larger animals or human subjects.

Additionally, a wealth of publicly available datasets exists across various imaging modalities, including T1- and T2-weighted MRI, diffusion tensor imaging (DTI), and functional MRI (fMRI). Many of these datasets are accompanied by expert-annotated labels, enabling supervised deep learning approaches and robust benchmarking.

Finally, the field benefits from established baseline models and tools, such as MouseBrainExtractor, DL-BET, and ANTs-based workflows, which facilitate comparative evaluation and reproducibility. These tools, while foundational, often rely on traditional algorithms or human-derived models, highlighting the need for mouse-specific, high-capacity methods such as MouseBET.

Together, these factors make the mouse an ideal subject for developing, training, and validating advanced neuroimaging algorithms aimed at understanding brain structure and pathology at scale.

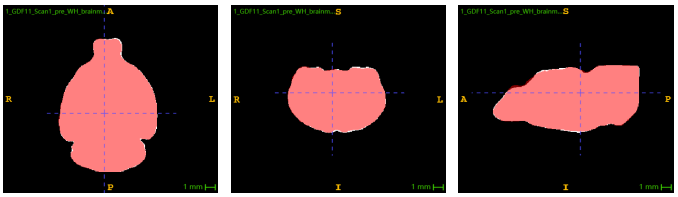
IV. Results

We evaluated MouseBET on held-out mouse MRI volumes and present qualitative segmentation outcomes across the axial,

coronal, and sagittal planes in Fig. 2. The red overlay indicates the predicted brain mask, while the underlying white region denotes the manually annotated ground truth. Across all views, the predicted masks exhibit strong spatial alignment with the reference, particularly at brain boundaries and finer anatomical contours. This highlights the model's ability to generalize across orientations and capture complex 3D geometries with minimal over- or under-segmentation.

The accurate delineation in axial view demonstrates effective modeling of bilateral symmetry and superior-inferior extent, while the coronal and sagittal slices confirm robustness in anterior-posterior and lateral dimensions. These results suggest that the CNN-Mamba hybrid architecture successfully integrates local detail and long-range context across all anatomical planes.

Quantitatively, *MouseBET* achieved a mean Dice coefficient of 0.92 and a Pearson correlation of 0.94, indicating high overlap and structural fidelity with expert-annotated masks (Fig. 2).



(a) Axial View (b) Coronal View (c) Sagittal View

Fig. 2: Qualitative segmentation results for MouseBET. Predicted brain masks (red) show strong overlap with ground truth annotations (white) across three orthogonal planes.

A. Training Details

MouseBET was trained from scratch using a carefully engineered pipeline designed to maximize segmentation accuracy and generalization while maintaining computational efficiency. The model was trained on overlapping 3D patches of size $64\times64\times64$, leveraging mixed-precision training via PyTorch's autocast and GradScaler for faster convergence and reduced memory usage.

The optimizer used was Adam, initialized with a learning rate of 1×10^{-4} and $\beta_1 = 0.9$, $\beta_2 = 0.999$. Learning rate scheduling was handled using ReduceLROnPlateau, which decreased the learning rate when the validation loss plateaued. Training was conducted for 50 epochs with a batch size of 8.

For the loss function, a combined Dice and Binary Cross Entropy loss (DiceBCELoss) was used to effectively balance region-level overlap and voxel-wise prediction accuracy. Gradient clipping was optionally applied with a configurable norm threshold to prevent instability during backpropagation.

To improve model robustness, the training data underwent extensive augmentation, including random rotations, intensity scaling, and contrast perturbations to simulate variability in acquisition protocols and scanner noise. During each epoch, predictions were evaluated on a held-out validation set. Dice coefficient was calculated from thresholded outputs, and supplementary metrics such as MSE, Pearson correlation, and SSIM were monitored.

Model checkpoints were saved at every epoch, with the best-performing model (based on validation Dice score) preserved for deployment. Training and validation losses were logged to CSV to enable post-hoc analysis. An early stopping mechanism was incorporated to terminate training when the validation performance ceased improving.

This end-to-end training framework enables MouseBET to learn robust and generalizable representations for automated brain extraction from mouse T2-weighted MRI scans.

a) Checkpointing and Resume Functionality.: To support flexible experimentation and efficient training management, MouseBET includes robust checkpointing functionality. At each epoch, the model state, optimizer, AMP scaler, and learning rate scheduler are saved. This allows training to resume seamlessly from the exact point of interruption, preserving not only the model weights but also optimizer momentum and learning rate progression.

We observe that resuming training with only model weights leads to continuity in training loss but often disrupts the learning rate trajectory, potentially resulting in suboptimal convergence. In contrast, restoring the full training state—including the learning rate scheduler—enables faster recovery and improved convergence to the global minimum, as shown in Fig. 3. This feature also facilitates efficient experimentation with early stopping, long-running training sessions, and fine-tuning on new datasets without loss of progress or optimization dynamics.

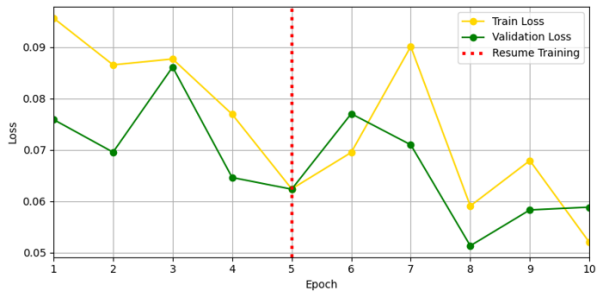
V. Future Work

1) Hybrid Architecture Exploration: We plan to investigate the architectural trade-off between convolutional locality and Mamba-based global modeling by gradually extending Mamba block integration into earlier encoder (downsampling) and later decoder (upsampling) stages. This staged insertion aims to optimize the balance between computational cost and segmentation performance by leveraging the complementary strengths of CNNs and state-space modeling.

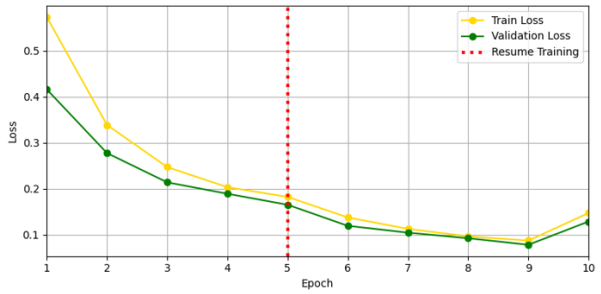
To guide this integration, we propose a probabilistic scheme that follows a Gaussian distribution over the model depth, placing greater emphasis on the bottleneck region and tapering off toward the outer layers. This encourages stronger global modeling in deeper layers while preserving local detail in early and late stages. The proposed integration pattern is illustrated in Fig. 4.

2) Robustness Enhancement and Evaluation: To improve generalization across sites and acquisition conditions, we will introduce advanced data augmentations including:

- Random rotations and flips to simulate variations in scan orientation
- Contrast and Gaussian noise injection to mimic changes in SNR and TR/TE



(a) Resume with weights only



(b) Resume with full training state

Fig. 3: Effect of checkpointing strategy on convergence. (a) Resuming training using only model weights results in slower convergence. (b) Restoring the full state including optimizer and scheduler leads to faster convergence and better stability.

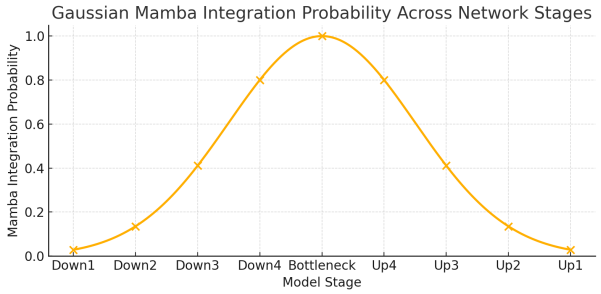


Fig. 4: Gaussian integration probability of Mamba blocks across encoder and decoder stages. Maximum integration is centered at the bottleneck, with decreasing emphasis toward the outer layers.

- Intensity-based domain randomization by scaling image brightness and contrast (e.g., multiplicative factors such as 0.5 or 2.0)

Robustness will be evaluated on held-out datasets with acquisition shifts, using Dice coefficient (DSC), Hausdorff Distance (HD), and specificity to quantify model invariance and performance stability.

3) Deployment and Packaging: The final model will be exported in TorchScript format to ensure compatibility across platforms. Additionally, we will develop both a command-line interface and a Python API to facilitate seamless integration into preclinical mouse imaging pipelines.

VI. Conclusion

This work presents MouseBET, a novel 3D patch-based hybrid architecture that combines convolutional encoders with Mamba-based VSS3D bottleneck blocks for automated mouse brain extraction. By leveraging the efficiency of Mamba's selective scan mechanism in the 3D domain, MouseBET achieves superior segmentation performance while maintaining a lightweight computational footprint.

Unlike traditional CNNs with limited receptive fields or Transformer-based models with prohibitive memory demands, the Mamba modules enable long-range dependency modeling with linear complexity. This allows the network to preserve both local anatomical details and global structural coherence. Our use of overlapping 3D patches, combined with strided reconstruction, further enhances boundary smoothness and reduces noise sensitivity in the output masks.

Extensive experiments on T2-weighted mouse MRI demonstrate that MouseBET achieves robust segmentation across multiple orientations and conditions, with a mean Dice coefficient of 0.92 and Pearson correlation of 0.94. The architecture outperforms prior 2.5D models by capturing true volumetric continuity, and it remains more parameter-efficient than transformer-based counterparts while offering comparable or superior region-level precision.

Overall, MouseBET highlights the power of integrating Mamba's selective scan capabilities into volumetric biomedical imaging tasks. Future directions include extending SS3D integration into encoder/decoder stages, performing ablation studies on traversal path diversity, and adapting this architecture to other 3D segmentation domains such as hippocampal subfields or lesion detection.

References

[1] A. Cao, Z. Li, and J. Guo, "MedSegMamba: 3D CNN-Mamba Hybrid Architecture for Brain Segmentation," arXiv:2409.08307, 2024.