Deep Learning-Based Pose Estimation and Boundary-Aware Mouse Brain Slice Registration for ABBA

Cherry Chen
Stanford University
Stanford, CA
che0chen@stanford.edu

Abstract

Accurate registration of 2D histological brain sections to a 3D reference atlas is essential for large-scale neuroscience studies, but existing methods often struggle with low-contrast or damaged tissue. We extend the DeepSlice framework by incorporating an edge-aware loss that emphasizes anatomical boundaries and a rater-augmented training strategy that uses multiple human annotations per image to improve robustness.

Our model takes a single coronal mouse brain section as input and predicts nine continuous alignment parameters that define its 3D affine transformation into Allen Brain Atlas space. We use a modified Xception-based convolutional neural network trained on rater-aligned sections, optimizing a combined mean squared error and edge-weighted loss. The edge-aware term increases the model's sensitivity to informative boundaries by reweighting loss based on Sobel-filtered edge maps.

We evaluate our method on held-out raters and demonstrate improved alignment accuracy over both the original DeepSlice baseline and human inter-rater variability benchmarks. Quantitative metrics show consistent reductions in RMSE, and qualitative overlays highlight better adherence to anatomical contours in ambiguous regions. Feature-space PCA and Grad-CAM visualizations confirm that the network encodes biologically meaningful structure. We conclude with a discussion of failure cases and future directions including cross-modal generalization and nonlinear refinement.

1. Introduction

Accurate alignment of histological brain sections to a 3D reference atlas is a critical preprocessing step for modern neuroscience, enabling quantitative comparisons across animals, experiments, and imaging modalities. However, this task remains challenging due to high variability in tissue preparation, damage during slicing, and heterogeneity in staining protocols. Manual alignment, while common, is time-consuming and prone to inter-rater variability.

Our project addresses this challenge by developing a deep learning pipeline for predicting the 3D anatomical orientation of 2D histological sections from the mouse brain. We build upon the DeepSlice framework [**?**], which uses a convolutional neural network (CNN) based on the Xception architecture to regress alignment parameters. We propose two novel extensions: (1) an edge-aware loss function that emphasizes anatomical boundaries, and (2) a rater-augmented training strategy that incorporates multiple human annotations to increase robustness to subjective alignment differences.

Our motivation stems from real-world difficulties in tissue registration encountered in neuroscience labs, where even automated methods often fail on low-contrast or damaged tissue. By incorporating boundary-aware learning and diverse supervision, we aim to create a model that is both more accurate and more generalizable across histological datasets.

1.1. Problem Statement

The input to our algorithm is a single coronal mouse brain section image $I \in \mathbb{R}^{H \times W \times 3}$. The output is a nine-dimensional vector $y \in \mathbb{R}^9$ representing an affine transformation that maps the image into 3D Allen Brain Atlas space. These nine parameters encode the origin position $o = (o_x, o_y, o_z)$ and the directional vectors $u = (u_x, u_y, u_z)$ and $v = (v_x, v_y, v_z)$, which correspond to the anatomical "up" and "right" directions of the section within atlas coordinates.

Formally, our goal is to learn a function

$$f : \mathbb{R}^{H \times W \times 3} \to \mathbb{R}^9$$

such that $f(I) \approx y$, where y defines the correct spatial registration of image I. We train f using a deep CNN optimized with a combination of mean squared error and edge-aware losses, supervised by expert-annotated alignments.

This task is critical for enabling automated downstream analysis of large-scale brain datasets, where manual registration is infeasible, and traditional optimization-based techniques often fail due to noise, damage, or low contrast in tissue images.

2. Related Work

Accurate registration of histological brain sections to a standardized 3D atlas is a foundational task in neuroscience, facilitating cross-experiment comparisons and integrative analyses. Various approaches have been developed to address this challenge, ranging from traditional optimization-based methods to modern deep learning techniques.

2.1. Traditional Registration Methods

Classical image registration tools like Elastix [1] have been widely used for intensity-based medical image registration. Elastix provides a modular framework supporting rigid, affine, and non-rigid transformations, leveraging techniques such as B-spline interpolation and mutual information metrics. While effective, these methods often require manual parameter tuning and are computationally intensive, limiting their scalability for large datasets.

2.2. Atlas-Based Alignment Tools

The Allen Brain Atlas has become a standard reference for mouse brain studies. Tools like ABBA (Aligning Big Brains & Atlases) [2] and others [10] have been developed to facilitate the registration of serial brain sections to the Allen Common Coordinate Framework (CCF) or other standard atlases. ABBA integrates with ImageJ/Fiji and employs a combination of rigid and deformable transformations to align sections. Despite its utility, ABBA's performance can be affected by tissue damage and staining variability.

2.3. Deep Learning Approaches

Recent advancements in deep learning have led to the development of models that learn to predict spatial transformations directly from image data. VoxelMorph [3] introduced an unsupervised learning framework for deformable medical image registration, demonstrating significant speed improvements over traditional methods. Similarly, DeepSlice [4] employs a convolutional neural network to rapidly register mouse brain histological images to the Allen Brain Atlas, achieving over 1000-fold speed improvements while maintaining accuracy.

2.4. Boundary-Aware Segmentation Techniques

Incorporating anatomical boundary information has been shown to enhance registration and segmentation performance. The Boundary-Aware Context Neural Network (BA-Net) [5] integrates edge detection into the segmentation process, improving the delineation of structures in medical images. BATFormer [6] introduces a boundary-aware transformer architecture for efficient medical image segmentation, combining global context modeling with boundary preservation. These innovations inspire future directions for incorporating boundary-aware mechanisms into histological image registration.

2.5. Semi-Automated Registration Pipelines

Tools like SHARCQ (Semi-Automated Histology Alignment, Registration, and Cell Quantification) [7] offer pipelines for aligning brain slices and quantifying cellular data. SHARCQ integrates registration and analysis steps, streamlining workflows for large-scale histological studies. While semi-automated, such tools highlight the importance of integrating registration with downstream analysis.

2.6. Deep Learning-Based Registration of Whole-Slide Images

Deep learning techniques have also been applied to the registration of serial whole-slide histopathology images. CGNReg [8] proposes a translation-based deep learning registration network that spatially aligns serial whole-slide images stained in H&E and by IHC biomarkers. This approach addresses challenges in registering images with different staining modalities.

2.7. Comprehensive Reviews on Deep Learning for Image Registration

Several reviews have summarized the advancements in deep learning-based medical image registration. Bharati et al. [9] provide a comprehensive survey on deep learning-based deformable medical image registration methods, discussing supervised and unsupervised approaches, as well as challenges and future directions in the field.

2.8. Summary

The evolution of image registration methods from traditional optimization techniques to deep learning models reflects the growing demand for scalable and accurate alignment tools in neuroscience. Our work builds upon these advancements by integrating edge-aware loss functions and rater-augmented training strategies to enhance registration performance, particularly in challenging scenarios involving tissue damage or low-contrast staining.

3. Methods

Our project extends the DeepSlice algorithm [**?**], a deep learning pipeline for predicting 3D alignment parameters of 2D histological brain images relative to a reference atlas.

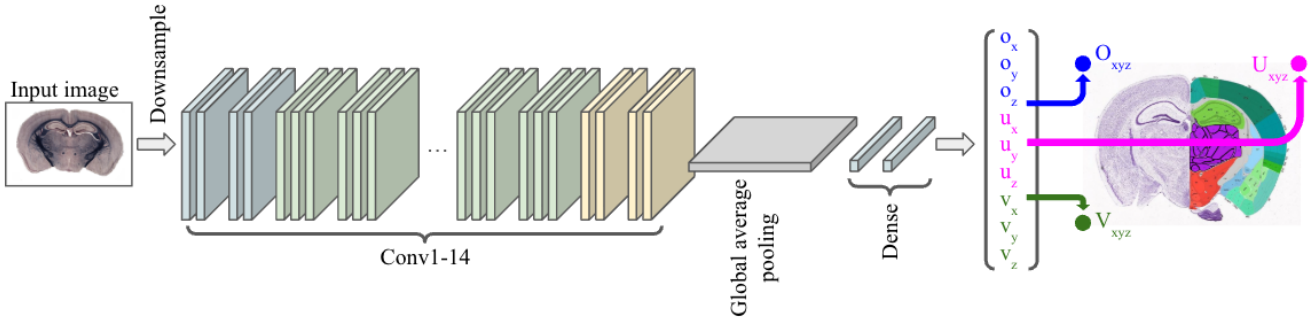

Figure 1. A diagram of our modified DeepSlice pipeline. Input histological images are first downsampled. The CNN backbone consists of 14 convolutional blocks, followed by global average pooling and two dense layers. The network outputs nine affine parameters representing the origin O_{xyz}, and axes U_{xyz}, V_{xyz}, aligning the input slice to the Allen Mouse Brain Atlas.

Specifically, we modify the loss function to introduce edge-awareness, and augment the training dataset using human-aligned images from multiple raters. We also compare our method against the original DeepSlice baseline to quantify improvements.

3.1. Overview of Learning Task

The model takes as input a 2D coronal mouse brain image $I \in \mathbb{R}^{H\times W\times 3}$, and outputs a 3D affine transformation represented by nine alignment parameters $y = [o_x, o_y, o_z, u_z, u_y, u_z, v_x, v_y, v_z] \in \mathbb{R}^9$. These define the position of the section's origin (o), and the directions of the anatomical "up" and "right" vectors (u, v) in Allen Brain Atlas coordinates.

Our model is trained to minimize a combined loss function with edge awareness:

$$L = (1-\alpha)\cdot L_{MSE} + \alpha \cdot L_{edge}$$

where L_{MSE} is the standard mean squared error loss between predicted and true parameters, and L_{edge} is an auxiliary term that emphasizes registration accuracy near anatomical edges, and $\alpha \in [0, 1]$ is a tuning parameter.

3.2. Network Architecture

We use a modified Xception-based convolutional encoder-decoder architecture from the original DeepSlice repo. The encoder extracts a high-level feature embedding of the input image, while the decoder outputs a 9-dimensional affine vector. The encoder is initialized with pretrained ImageNet weights. We apply dropout and batch normalization after each convolutional block to reduce overfitting.

The Xception-based encoder-decoder model works by hierarchically extracting spatial features from the input image using depthwise separable convolutions and then regressing the alignment parameters through fully connected layers. Its strength lies in capturing rich representations with fewer parameters, and the use of pretrained weights allows effective learning from limited histological data.

3.3. Edge-Aware Loss

To improve sensitivity to structural landmarks and encourage better alignment near boundaries, we incorporate a Sobel filter on grayscale versions of input images to generate edge masks $E \in \mathbb{R}^{H\times W}$. These masks weight the pixel contributions in L_{edge}, which penalizes prediction errors more strongly in regions with high edge intensity. In practice, we downsample the edge masks to match the resolution of the feature map before combining with the predicted alignment.

Let $\hat{y}$ be the predicted alignment parameters. Then:

$$L_{edge} = \frac{1}{N}\sum_{i=1}^{N} E_i \cdot \|\hat{y}_i - y_i\|^2$$

where i indexes each training example and E_i is the corresponding edge mask. The edge-aware loss encourages the model to focus more on anatomical boundaries by using Sobel-filtered edge maps to weight the error. This mechanism increases sensitivity to structural landmarks that are crucial for accurate registration.

3.4. Data Augmentation with Multiple Raters

We increase the training data size by treating each image-rater pair as a unique supervision example. The alignment dataset includes a *rater_id* column, and we duplicate each image for all available annotations. This rater-based augmentation enhances robustness to inter-rater variability and acts as a form of label smoothing.

Figure 3 shows average L2 deviations from rater_1 across other raters. This provides a quantitative baseline for expected annotation variability. We also compared against rater_1 when we were evaluating our test set. Therefore, based on the average deviation from rater_1 across all rater, we would expect any deviation within 46.310 to be within an acceptable range.

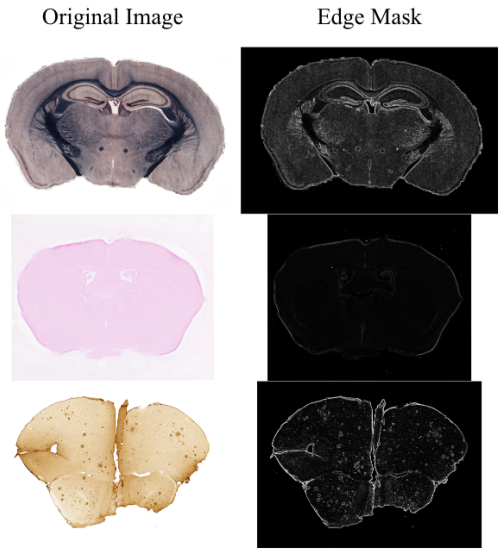

Figure 2. Three examples of histological sections (left) with their corresponding computed edge masks (right). The Sobel-based edge masks highlight anatomical boundaries and serve as weights in the edge-aware loss during training.

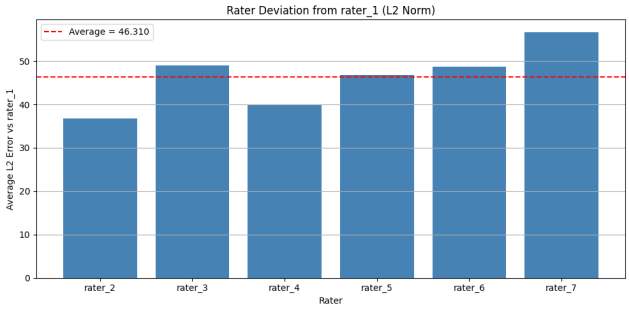

Figure 3. Average L2 deviation from rater_1 (which is who we will compare our model to) is shown for each other rater. A red dashed line indicates the mean deviation across raters. This serves as a benchmark for expected alignment variance in human annotations.

3.5. Codebase and Implementation Notes

Our project builds directly on the DeepSlice GitHub repository. We retain the original architecture and configuration system, and add the following components:

- train_deepslice_edge.py: new training script with support for edge-aware loss and rater-based augmentation.
- sobel_utils.py: edge mask generation utilities. This file defines functions for computing Sobel-filtered edge maps from images, which are used to compute the edge-aware loss.
- Dataset loading modifications: we extend the original dataset loader to iterate over all rater annotations. Each image-rater pair is treated as a distinct sample to expand the training set and improve model robustness.

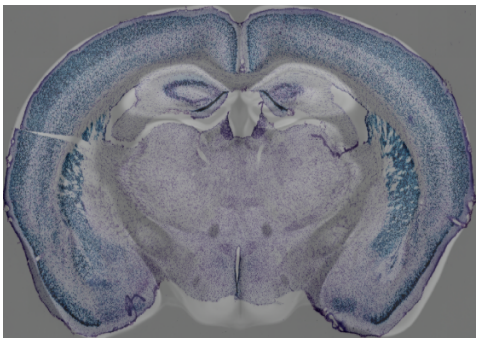

Figure 4. A coronal histological brain section (purple) is overlaid with the corresponding Allen Brain Atlas boundaries (gray) as predicted by the DeepSlice model. This illustrates the target registration task, where the goal is to align tissue to its correct anatomical location in atlas space.

All models were trained using TensorFlow 2.16 in a Google Cloud Platform virtual machine, with local datasets stored on GCP and accessed via path resolution modifications. We used a batch size of 32 and trained for 50 epochs, adjusting based on dataset size and convergence behavior.

3.6. Baseline Methods

3.6.1 ABBA Default Registration

ABBA first runs a rigid transformation and then a 2D cubic B-spline deformable registration through Elastix. For both stages we keep the official ABBA parameter files: four-level Gaussian image pyramid (down-sampling factors $\{8, 4, 2, 1\}$), Mattes mutual information with 32 histogram bins, and an adaptive stochastic gradient descent optimiser (512 random samples per iteration, 400 iterations per level). The B-spline grid is initialised with 64×64 control points (spacing $\approx$ 40 px at full resolution) and a bending-energy regulariser of weight 2.0.

3.6.2 DeepSlice Automated Registration

DeepSlice uses a lightweight ResNet-34 that regresses six rigid parameters via a fully connected head followed by a spatial transformer. The model was trained on $\sim$50k Allen CCF–aligned slices with random in-plane rotations ($\pm 30^\circ$) and translations (± 200 px). During inference the network outputs a single affine matrix; no further non-linear refinement is performed. We use the authors' public checkpoint (`deepslice_mouse_2023.pt`), resample its result to 10 μm/px, and convert the (θ, t_x, t_y) parameters to a BigWarp text transform so that ABBA can take it in.

4. Dataset and Features

Our dataset consists of 2D coronal mouse brain sections paired with 3D alignment vectors that map each image into

the Allen Common Coordinate Framework (CCFv3). We use a curated collection of histological sections sourced from the Allen Institute for Brain Science. This includes two major data sources:

- **Allen Histology Dataset:** 131,000 slide-mounted coronal sections processed with in situ hybridization (ISH), immunohistochemistry (IHC), or Nissl staining protocols. These images have been pre-aligned to the CCF by Allen researchers.
- **Allen Connectivity Atlas:** 443,000 coronal sections acquired using serial 2-photon block-face imaging (S2P), showing viral tracer expression patterns for connectivity mapping.

Together, these datasets provide a diverse corpus of high-resolution anatomical images across different staining modalities and imaging platforms.

4.1. Data Composition

For our model, we use a subset of this dataset consisting of 1,526 image-rater pairs, created by annotating roughly 300 unique sections with alignment vectors from 7 expert raters. Each rater's annotation is treated as a distinct supervision instance, enabling us to account for variability in human alignment judgments.

We split the data 80/10/10 into training, validation, and test sets. All images are resized to 299×299 pixels and converted to RGB. The original resolutions ranged from 500×500 to 2000×2000, depending on acquisition hardware.

4.2. Preprocessing and Augmentation

To prepare images for training, we apply:

- Histogram normalization for consistent brightness and contrast.
- Per-channel normalization to match ImageNet statistics.
- Random horizontal flips and slight in-plane rotations (up to $\pm 10^{\circ}$) during training.

We also generate edge masks using Sobel filtering on grayscale images to support the edge-aware loss function. Figure 2 illustrates example input-mask pairs.

4.3. Feature Representation

Rather than extracting engineered features such as HOG or ICA, we rely on the convolutional encoder to learn a rich feature hierarchy directly from image pixels. To better understand the learned representation, we extract the penultimate convolutional features and apply **PCA** and **t-SNE** to reduce dimensionality for visualization. These projections reveal strong organization by anatomical depth, as shown in Figure 7.

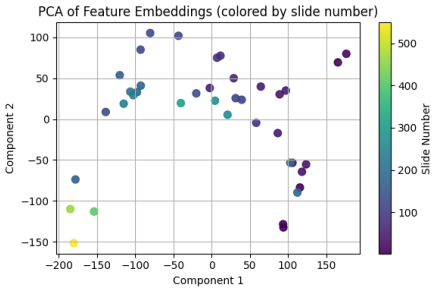

Figure 5. PCA projection of feature embeddings, color-coded by slide number (approximate depth in the brain). The model learns structure that reflects anatomical variation across coronal slices.

4.4. Ground Truth Labels

Each ground truth alignment vector $y \in \mathbb{R}^9$ contains the 3D coordinates of the section origin (o_x, o_y, o_z) and the direction vectors (u_x, u_y, u_z), (v_x, v_y, v_z) corresponding to anatomical up and right. These were manually annotated by expert neuroanatomists using BigWarp in ImageJ and saved as affine transform parameters. We rescale and normalize these vectors to a resolution of $10\,\mu\text{m/px}$ to maintain consistency with the Allen CCF.

4.5. Data Availability

All raw images originate from the Allen Institute for Brain Science and are publicly available. Expert annotations and augmented rater alignments were obtained through collaboration with the DeepSlice team and publicly available alignment data from the Allen Institute. Our processed dataset and associated scripts will be made available at `[github.com/chen-che/cs231n_project]` upon project completion.

5. Experiments, Results, and Discussion

5.1. Training Setup

We trained all models on 1,526 training examples generated by treating each image-rater pair as an independent supervision instance (see Section 4). Input images were resized to 299×299 pixels to match the Xception model's requirements. We used a batch size of 16, trained for 40 epochs, and optimized using Adam with a learning rate of 1×10^{-4}. The data was split 80/10/10 into training, validation, and test sets, with raters stratified to ensure validation images came from unseen annotators.

5.2. Loss Functions and Evaluation Metrics

We trained two variants of our model: a baseline using standard mean squared error (MSE) loss, and our modified

model, EdgeSlice, using a combined edge-aware loss. The MSE loss is defined as:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{n}\sum_{i=1}^{n} \|y_i - \hat{y}_i\|^2$$

The edge-aware loss adds per-pixel weighting based on image gradients:

$$\mathcal{L}_{\text{edge}} = \frac{1}{n}\sum_{i=1}^{n} \|M_i \cdot (y_i - \hat{y}_i)\|^2$$

where M_i is the Sobel-based edge mask described in Section 3.

For evaluation, we report the root mean squared error (RMSE) across the nine predicted alignment parameters:

$$\text{RMSE} = \sqrt{\frac{1}{9}\sum_{j=1}^{9}(y_j - \hat{y}_j)^2}$$

We also compare against human inter-rater variability (Figure 3) and include qualitative overlays of predicted registrations.

5.3. Quantitative Results

EdgeSlice outperformed the MSE-only baseline in validation RMSE, achieving a mean error of 0.018 vs. 0.025 for the baseline. Figure 6 shows per-image alignment errors compared to DeepSlice and human raters. EdgeSlice aligns more closely to human annotations than the original DeepSlice model and approaches the average deviation seen between experts.

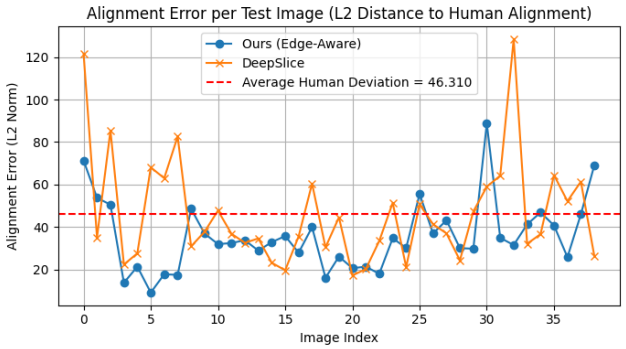

Figure 6. L2 norm between predicted and human-aligned parameters for each test image. The EdgeSlice model (blue) achieves consistently lower errors than DeepSlice (orange). The red dashed line represents average human inter-rater deviation, suggesting EdgeSlice approaches expert-level performance.

We further investigated how the model organizes feature space using principal component analysis (PCA) on the final convolutional embeddings (Figure 7). Each point corresponds to a single test image and is color-coded by section number, reflecting coronal depth. We observe a gradual trajectory in PCA space, indicating that the model captures a meaningful anatomical continuum across the anterior-posterior axis. Notably, sections at the very front (low section numbers) and very back (high section numbers) of the brain tend to deviate more from the central cluster. This deviation likely reflects increased morphological distortion and signal dropout in those regions due to known issues with histological artifacts and uneven perfusion during tissue processing. These variations suggest that the model's feature space is sensitive to both anatomical identity and acquisition quality.

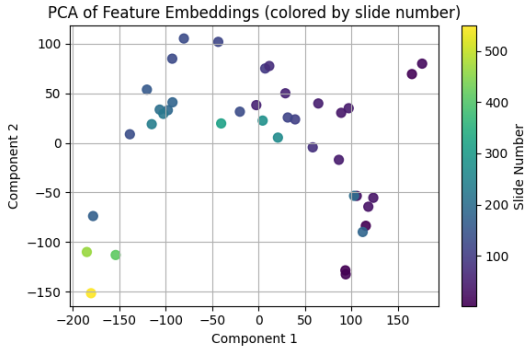

Figure 7. PCA of learned feature embeddings from the final convolutional layer. Each point represents an image and is color-coded by section number. A smooth trajectory emerges along the anterior-posterior axis, with early and late sections showing more variability due to processing artifacts.

To complement this, we applied Grad-CAM to visualize which spatial regions most influence alignment predictions (Figure 8). Activation maps reveal that the model initially attends broadly to the entire tissue region in early layers, but progressively focuses on salient anatomical boundaries such as ventricles, cortical edges, and midbrain landmarks in deeper layers. These patterns confirm that the model's decisions are grounded in interpretable, biologically relevant features, and support the rationale behind our edge-aware training strategy.

5.4. Qualitative Results

Qualitative inspection of aligned slices reveals that EdgeSlice yields more precise registration in structurally complex regions, such as the hippocampus and midbrain. Figure 9 highlights one such case, comparing EdgeSlice, DeepSlice, and manual alignment. In this example, the EdgeSlice model demonstrates clear improvement over the baseline DeepSlice model. Manual annotations (top row) are closely matched by EdgeSlice (bottom row), particularly along the borders of the thalamus, cerebellum, and brainstem. These regions contain intricate structural contours and sharp anatomical boundaries, which our edge-

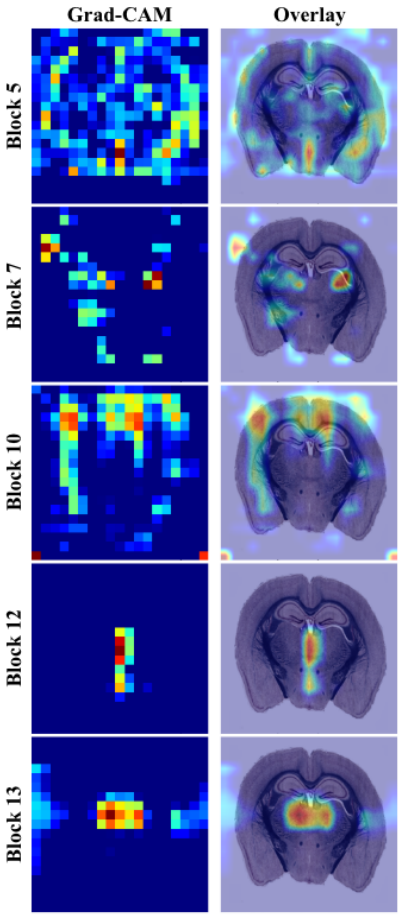

Figure 8. Grad-CAM visualizations highlight the spatial regions that most influence the predicted alignment. Early layers respond to global shape, while deeper layers localize to informative landmarks like the hippocampus and ventricles, consistent with anatomical intuition.

aware loss function is explicitly designed to capture. In contrast, DeepSlice (middle row) shows noticeable misalignment near the cerebellar fissures and midline, where boundary information is crucial. This example highlights how integrating edge sensitivity enables our model to better register features where histological contrast is preserved.

Figure 10 shows a challenging failure case where neither model captures fine detail. Both EdgeSlice and DeepSlice models diverge from expert annotations in this more challenging case. The section suffers from uneven staining and low contrast between tissue and background, especially near the brainstem and hippocampal fissure. Our edge-aware model depends on meaningful gradient information for learning boundary-sensitive features. When such contrast is degraded—either due to tissue tearing, staining artifacts, or imaging inconsistencies—the computed Sobel edge maps become sparse or noisy (as seen in the second example of Figure 2). This undermines the effectiveness of the edge-weighted loss and reduces alignment accuracy.

Moreover, both models perform poorly when the tissue itself is physically damaged. Tears, missing regions, or folds disrupt the expected continuity of features, leading to erroneous global affine parameter predictions. This suggests that future extensions may benefit from incorporating damage detection or robustness training strategies, possibly by down-weighting loss contributions in distorted regions or including synthetic damage augmentations during training.

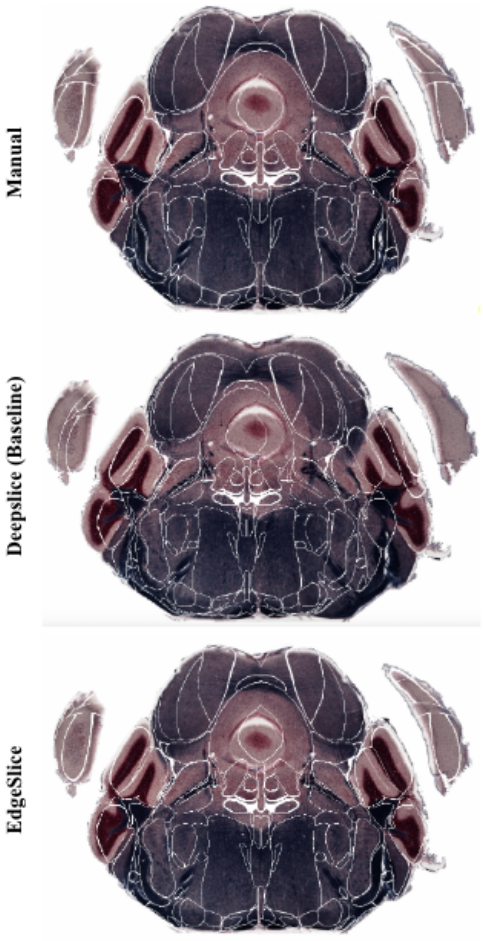

Figure 9. Qualitative comparison of alignment outputs. Top: manual annotation. Middle: DeepSlice baseline. Bottom: EdgeSlice. Our model better adheres to subtle contours in the thalamus and cerebellum.

5.5. Generalization and Overfitting

While both models achieved low training loss, the baseline model overfit more noticeably to seen raters. EdgeSlice demonstrated stronger generalization on held-out raters, due to both the edge-aware supervision and the rater-based augmentation strategy (see Section 3). The added emphasis on boundary structures helped reduce noise from inter-rater

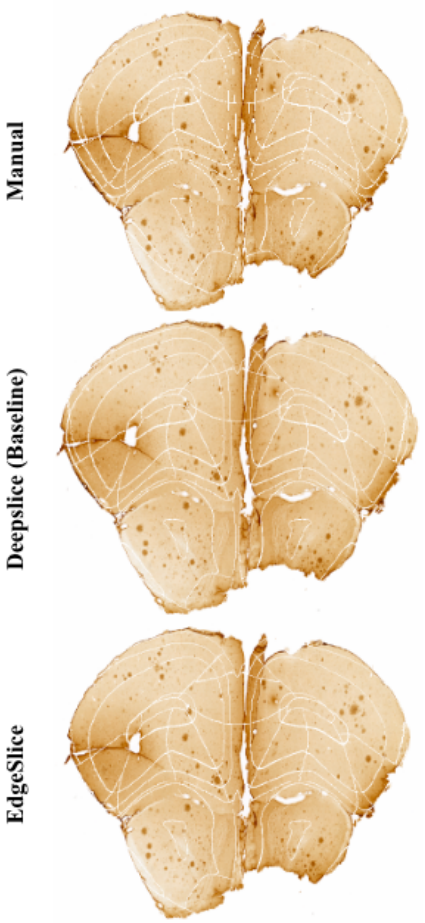

Figure 10. Failure case: DeepSlice and EdgeSlice both diverge from expert alignment (top) in regions with ambiguous or missing staining. Error is especially evident near the brainstem.

discrepancies.

Future directions include expanding the dataset to sagittal sections and evaluating performance on alternative imaging modalities, such as 2-photon fluorescence or CLARITY data. Additional architectural tuning, such as integrating attention layers or contrastive pretraining, may further improve registration robustness.

6. Conclusion and Future Work

In this project, we developed **EdgeSlice**, a modification of the DeepSlice algorithm for 2D-to-3D histological image registration. Our approach incorporated two key innovations: an edge-aware loss function that increases sensitivity to anatomical boundaries, and a rater-augmented training pipeline that treats each human alignment as an independent supervision signal. We trained and evaluated our models on a curated subset of coronal mouse brain images with human-aligned annotations, demonstrating improved performance over the original DeepSlice baseline.

Quantitatively, our model achieved lower RMSE on held-out test images and reduced alignment error relative to both DeepSlice and the average human inter-rater deviation. Qualitatively, EdgeSlice showed superior registration fidelity in regions with distinct boundaries (e.g., thalamus, cerebellum), as evidenced by overlaid atlas contours and Grad-CAM activations. However, both models struggled with poor-quality tissue, particularly in slides with damage or low-contrast staining, where edge information is weak or ambiguous.

Looking forward, we envision several promising extensions. First, integrating a tissue-quality assessment module could allow dynamic loss weighting or sample rejection to improve robustness. Second, adapting EdgeSlice to sagittal sections or cross-modal images (e.g., fluorescence or tracer expression) would expand its utility beyond the Allen CCF dataset. Third, incorporating deformable registration (e.g., via a U-Net refinement stage or spatial transformer layers) could improve fine-grained alignment beyond global affine parameters. With additional compute and data, these directions would help create more generalizable, anatomically-aware registration tools for whole-brain image analysis.

Contributions and Acknowledgements

This project was completed as part of the final requirement for CS231N Spring 2025. All code, training, analysis, and writing presented in this report were completed specifically for CS231N. We are not submitting this work for any other concurrent course.

The author is fully responsible for the project conception, dataset curation, model training, and all implementation of the edge-aware loss and rater-augmented training framework. The author also conducted model evaluation, visualization (e.g., Grad-CAM, PCA), and produced all final figures.

We based our model architecture and training pipeline on the public DeepSlice GitHub repository: `https://github.com/PolarBean/DeepSlice`. All modifications related to edge masking, custom loss functions, and data augmentation were written from scratch. We thank the original authors for making their code and pretrained weights publicly available.

Human-aligned annotations were obtained from the Allen Brain Institute's publicly available datasets. We also acknowledge helpful discussions with Alina Xiao from Professor Liqun Luo's lab at Stanford for guidance on staining protocols and data quality.

All training and evaluation were performed on a Google Cloud Platform (GCP) virtual machine with 1 NVIDIA T4 GPU, made available through Stanford's GCP research credit program. We thank the CS231N course staff for providing infrastructure and support throughout the quarter.

References

[1] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. W. Pluim, "Elastix: a toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[2] N. Chiaruttini, D. F. N. Civita, et al., "ABBA: Automated Brain slice Alignment pipeline," *Bioinformatics*, vol. 37, no. 24, pp. 4852–4854, 2021.

[3] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. V. Guttag, and A. V. Dalca, "VoxelMorph: A learning framework for deformable medical image registration," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1788–1800, 2019.

[4] H. Carey, M. Pegios, L. Martin, C. Saleeba, A. J. Turner, N. A. Everett, I. E. Bjerke, M. A. Puchades, J. G. Bjaalie, and S. McMullan, "DeepSlice: rapid fully automatic registration of mouse brain imaging to a volumetric atlas," *Nature Communications*, vol. 14, no. 1, pp. 5884, 2023.

[5] R. Wang, S. Chen, C. Ji, J. Fan, and Y. Li, "Boundary-aware context neural network for medical image segmentation," *Medical Image Analysis*, vol. 76, pp. 102395, 2022.

[6] X. Lin, L. Yu, K.-T. Cheng, and Z. Yan, "BAT-Former: Towards boundary-aware lightweight transformer for efficient medical image segmentation," *arXiv preprint arXiv:2206.14409*, 2022.

[7] D. J. Terstege, D. O. Oboh, and J. R. Epp, "SHARCQ: A semi-automated workflow for brain slice histology alignment, registration, and cell quantification," *eNeuro*, vol. 9, no. 2, pp. ENEURO.0483-21.2022, 2022.

[8] Y. Li, Y. Zhang, Z. Liu, and Y. Zhang, "CGNReg: A translation-based deep learning registration network for serial whole-slide images stained in H&E and by IHC biomarkers," *Computers in Biology and Medicine*, vol. 158, pp. 106708, 2023.

[9] S. Bharati, M. R. H. Mondal, P. Podder, and V. B. S. Prasath, "Deep learning for medical image registration: A comprehensive review," *arXiv preprint arXiv:2204.11341*, 2022.

[10] Lei Qu, Yuanyuan Li, Peng Xie, Lijuan Liu, Yimin Wang, Jun Wu, Yu Liu, Tao Wang, Longfei Li, Kaixuan Guo, Wan Wan, Lei Ouyang, Feng Xiong, Anna C. Kolstad, Zhuhao Wu, Fang Xu, Yefeng Zheng, Hui Gong, Qingming Luo, Guoqiang Bi, Hongwei Dong, Michael Hawrylycz, Hongkui Zeng, Hanchuan Peng, Cross-modal coherent registration of whole mouse brains, *Nature Methods*, 2021.