Supplementary Material

1 Requirements In Data Preparation

The extraction of positional information in our approach is dependent on the co-registration of high-resolution brain sections. This process involves aligning a sequence of brain sections with respect to a designated reference section, ensuring consistency across the stack. The selection of the initial reference section plays a pivotal role in determining the orientation and accuracy of the entire aligned stack.

The orientation of the reference section directly influences the alignment of other sections, as the initial alignment will propagate through the stack. Therefore, careful consideration of the reference section's alignment in the 2D plane is crucial. The initial reference selected should hold as an appropriate choice to cater for both one/ two hemisphere brain image data.

In our analysis, as illustrated in Figure 1, we determined that the third section in the sequence is the most suitable choice for the initial reference. This section exhibits an orientation that aligns with the conventionally considered biological vertical axis, ensuring anatomical consistency. Such alignment is critical for accurately capturing positional relationships and minimizing distortions during co-registration.

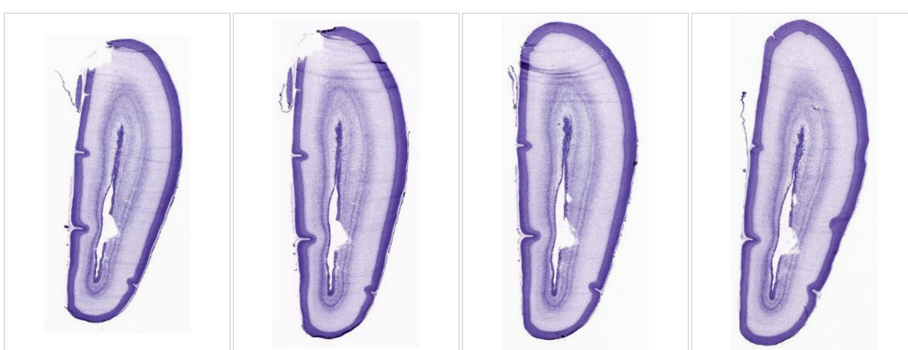

Figure 1: A sequence of brain sections from Allen 21 pcw brain, highlighting the misalignment present within the brain sections.

2 Table Of Notations

Table 1: Table of notations considered for $PosDiffAE$ model.

Notation	Definition
X, H, W	High-resolution (HR) image, height and width of the HR
x, h, w	A patch from the HR image, height and width of the patch
A	Ground truth annotation
a_{reg}	Region label assigned to patch x_0 by the classifier
$Enc_{\Phi'}$	Learnable encoder integrated into the DiffAE structure
f_{Φ}	UNet structure of DiffAE
z_{sem}	Latent space embedding generated by $Enc_{\Phi'}$
f_{dim}	Dimensionality of z_{sem}
N	Standard Gaussian distribution
T	Total time steps considered for the diffusion process
t	Intermediate time representation.
x_t	Noisy version of the image x_0 at time t
x_{t-1}	Noisy version of the image x_0 at time $t-1$
x_T	Image sampled from the standard Gaussian distribution at time step T
β_t	Hyper-parameters representing noise levels at time step t
α_t	Cumulative β_t
ϵ_t	Noise estimated by the model at time t
$p_\phi(x_{t-1}\|x_t, z_{sem})$	Distribution learnt by the model
$q(x_{t-1}\|x_t, x_0)$	Conditional prior distribution
L_{MSE}	Mean squared error loss
c_x, c_y	Centroid coordinates of x
x_p, y_p	Top-left corner coordinates of the patch within x
r_X	Maximum value of r_0 across all patches in a section X
r_0	Normalized radial distance between (x_p, y_p) and (c_x, c_y)
θ	Angle between r_0 and the horizontal line passing through (c_x, c_y)
α	Alignment (rotation) angle between brain sections, computed during inter-section registration
θ_0	Total radial angle
$f_r, f_{\theta'}$	Fully connected linear layers used for regressing distance and angle
r_p	Radial distance regressed by f_r
θ_p	Radial angle regressed by $f_{\theta'}$
L_r	Radial distance loss
L_θ	Radial angle loss
L_{total}	Total loss function
$\lambda_1, \lambda_2, \lambda_3$	Hyperparameters for weighting L_{MSE}, L_r, L_{theta}

Table 2: Table of notations considered for tear artifact restoration utilizing $PosDiffAE$ model.

Notation	Definition
M	Tear mask with indicating tear regions
x_{ROI}	ROI from tear region
a_{ROI}	Ground truth annotation for x_{ROI}
m_{ROI}	Tear mask for x_{ROI}
$x_{r,c}$	Image patch at row r and column c within x_{ROI}
$m_{r,c}$	Tear mask for $x_{r,c}$
z'_{sem}	Interpolated conditioning vector
$x^{r,c}_{noisy}$	Noisy version of patch $x_{r,c}$ sampled from Gaussian distribution
$\hat{x}^{t-1}_{r,c}$	Intermediate restored image generation for $x_{r,c}$
$\tilde{x}^{t-1}_{r,c}$	Final restored image generation for $x_{r,c}$
$\tilde{x}_{ROI}$	Final restored image for x_{ROI} after stitching
$\tilde{X}$	Final artifact-free restored HR image
$N^{'}$	Number of denoising steps adaptively chosen to recover tear

Table 3: Table of notations considered for tear JPEG restoration utilizing $PosDiffAE$ model.

Notation	Definition
x_{compr}	JPEG compressed image.
$T^{'}$	Noising time step, adaptively chosen based on the Quality Factor (QF) of x_{compr}
z^{compr}_{sem}	Latent semantic embedding of x_{compr}
$x^{compr}_{T'}$	Noisy version of x_{compr} at the noising time step $T^{'}$
$\alpha_{T'}$	Parameter representing the noise scale in the Gaussian diffusion process at time step $T^{'}$
$\epsilon_{T'}$	Gaussian noise sampled from the standard normal distribution at $T^{'}$
$N^{''}$	Number of adaptive denoising steps used to recover the x_{compr}

3 Hyper-parameter Selection

The hyper-parameter selection process in our method is detailed below:

- *Training hyper-parameters*- We have chosen these hyper-parameters based on the accuracy of region classification. The details of training hyper-parameters are detailed below in Table 4.

- *Inference hyper-parameters*- We have chosen these parameters based on the performance of tear artifact and JPEG restoration. *Tear Artifact Restoration*- The denoising parameter $N^{'}$ was set to 50, as this denoising time step best satisfies the trade-off between cellular distribution in the restored image set and complexity involved during denoising. *JPEG Restoration*- The choice of noising and denoising parameters depended on the JPEG Quality Factor (QF). Since a lower QF (QF = 5) implies higher degradation, more noising and denoising ($T^{'}$=20, $N^{''}$=50) steps were necessary and for QF = 10 ($T^{'}$=20, $N^{''}$=40) and QF = 15 ($T^{'}$=10, $N^{''}$=40), where more structural information remained intact, fewer steps were required.

Table 4: Table highlighting the variation of classification accuracy for different hyper-parameter values.

Hyper-parameters	**Variation in values**	**Classification Accuracy (%)**
Regression Loss Component (λ_2)	0.1/ 0.01/ **0.001**	60/ 65/ **78**
Regression Loss Component (λ_3)	0.1/ 0.01/ **0.001**	59/ 67/ **78**
Positional Encodings	**Radial Distance & angle**/ Spatial coordinates	**78**/ 76
Layers in MLP for regression	**3**/ 10/ 20	**78**/ 78/ 77
Latent Dimensions	128/ **512**/ 1024	69/ **78**/ **78**

4 Statistical Significance

The discussions related to statistical significance are as follows:

- *Cohen's Kappa (κ)*- Our method achieved the highest κ among all compared methods, aligning with the method-wise ranking of classification accuracy. This alignment arises because the agreement between the model's predictions and ground truth is used to calculate κ.
- *Paired t-test*- A paired t-test comparing the MSE distributions of our model and the baselines yielded $p < 10^{-7}$, indicating a statistically significant difference.
- *Wilcoxon Rank Sum Test*- A Wilcoxon rank sum test comparing the PSNR/SSIM distributions of our model against the baselines resulted in $p < 10^{-3}$, confirming the statistical significance of the observed differences.