REPRESENTATIONAL ALIGNMENT OF GLOMERULI ACTIVATION IN MURINE OLFACTORY BULB

Vivek Kumar Agarwal*
Center for Data Science
Center for Neuroscience
New York University
New York, USA
vka244@nyu.edu

Mario Garriado
Center for Data Science
New York University
New York, USA
mario.garrido@nyu.edu

Julia Manasson
Center for Data Science
New York University
New York, USA
manasj02@nyu.edu

Ilia Sucholutsky
Center for Data Science
New York University
New York, USA
is3060@nyu.edu

ABSTRACT

The presence of disease results in the production of odors different from the healthy state, which can be detected by mice. Each odorant generates a unique glomeruli activation pattern in the murine olfactory system. However, these signals differ spatially and temporally across mice and capturing them is labor-intensive. To optimize the process, glomeruli activation patterns for many odorants could be mapped from a reference animal imaged on all available odorants to target animals imaged on a subset of odorants. In our initial approach to develop this method, we align the glomeruli activation patterns for 11 odorants between 2 mice using the optimal transport Sinkhorn algorithm.

1 INTRODUCTION

The presence of disease alters host cell metabolism, resulting in the production of abnormal biochemical profiles and the generation of odors atypical of healthy cells (Piqueret et al., 2023). Changes in odor are often subtle and undetected by human olfaction, but can be detected by animals with a keen sense of smell (Bijland et al., 2013). Several examples include dogs detecting lung cancer in breath samples (Bijland et al., 2013), rats detecting Mycobacterium tuberculosis in sputum (Bijland et al., 2013), and mice detecting bladder cancer (Sato et al., 2017) and prostate cancer (Sato et al., 2022) in urine. If sensitive enough, this approach could have significant impact on noninvasive diagnostics and disease classification, with applications that range from malignancy screening to discrimination between bacterial and viral infections, enabling optimal antibiotic use.

Biospecimens from patients with disease produce odors comprising complex mixtures of one or more odorants (chemical stimuli capable of evoking a smell) (Gottfried, 2010). In mice, each odorant produces a unique neuronal signal by activating a pattern of glomeruli (structures in the olfactory bulb that initially process olfactory stimuli) (Soucy et al., 2009; Burton et al., 2022). These glomeruli activation patterns, in turn, can be used as biomarkers for the presence of disease. However, the same odorant can exhibit spatial and temporal variation in glomeruli activation patterns across mice. Furthermore, capturing glomeruli activation patterns is costly, resource-heavy, and time consuming, requiring the generation of a cranial window to capture neuronal activation signals and specialized training (Agarwal et al., 2024; Yeon, 2022). To optimize this process, we envision a likely solution in imaging a small number of reference mice across the complete set of odorants and a larger cohort of target mice on a subset of odorants. The remaining odorants can then be mapped from the reference mice to the target mice. As a foundational step, we aligned the glomeruli activation patterns of odorants between two mice and evaluated the success of this alignment.

*For queries on paper/data/code, please reach out via email: vka244@nyu.edu.

2 DATA

2.1 DATASET

We used the publicly available oMNIST dataset, which comprises videos of glomeruli activation patterns from 2 mice exposed to 65 odorants (Agarwal et al., 2024). To explore the feasibility of olfactory representational alignment, we focused on a subset of 2 mice exposed to 11 odorants with 10 trials for each odorant (Agarwal et al., 2024). Each mouse was exposed to odorants at specific concentrations through an olfactometer and the florescence in the olfactory bulb was recorded via a cranial window. In total, 220 video files (2 mice x 11 odorants x 10 trials) were utilized for the analysis.

2.2 PREPROCESSING

Videos were converted into tiff stacks of 320 image frames at 256 x 256 resolution. Glomeruli activation patterns in the raw videos were initially masked by the vasculature, which required extensive preprocessing (Fig 1). Background subtraction (Gibson & Bovik, 2000) followed by denoising with Anisotropic diffusion (Weickert, 1996) with a convolutional kernel were performed to extract neuronal signals from the noisy images. These techniques were applied to each of the 320 image frames in the 220 video files. The 320 image frames were subsequently compressed into 1 frame to create a single maximum pixel intensity (MPI) image. Although the temporal component was lost with this approach, it allowed for a more feasible method of aligning and analyzing the spatial components of glomeruli activation patterns.

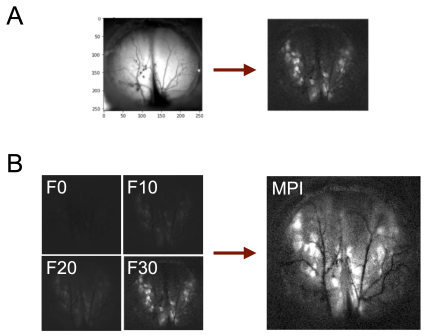

Figure 1: Data preprocessing. A) Image preprocessing to uncover glomeruli activation patterns masked by vasculature. B) Image frames from each trial were compressed into a single MPI frame.

3 METHODS

Optimal transport was used to spatially align odorant glomeruli activation patterns between two mice; representations learned from the reference mouse (1952) were aligned to that of the target mouse (1953). Prior to alignment, non-negative matrix factorization (NNMF), Segment Anything Model 2 (SAM2), and Autoencoder were used to extract representations from the MPI images.

3.1 NON-NEGATIVE MATRIX FACTORIZATION (NNMF)

NNMF is an unsupervised learning algorithm that factorizes a non-negative target matrix V into the product of two lower rank non-negative matrices W and H that approximate the original matrix such that:

$$V \approx W \cdot H, \text{where } (W, H) \geq 0.$$

This reduces the dimensionality of the original matrix (Lee & Seung, 1999). We relied on NNMF as a simple base case for extracting MPI image representations.

3.2 SEGMENT ANYTHING MODEL 2 (SAM2)

SAM2 is a robust tool used for image segmentation (Ravi et al., 2024). SAM2's Hiera vision encoder (Ryali et al., 2023) has demonstrated strong capabilities in learning hierarchical visual features, making it particularly suitable for capturing the complex spatial patterns present in glomeruli activation pattern images. The Hiera encoder follows a hierarchical design with four stages of progressively decreasing spatial resolution but increasing channel capacity. This multi-scale architecture allows the model to efficiently capture both fine-grained local features in early stages and more abstract global patterns in later stages. The encoder processes each MPI image $\mathbf{x}_t$ to produce feature embeddings $\mathbf{F}t = \mathcal{E}\text{img}(\mathbf{x}t)$, where $\mathcal{E}\text{img}$ represents the Hiera encoder.

We leveraged a medical-domain-adapted version of SAM2 (Zhu et al., 2024) that was pre-trained on a large corpus of medical imaging data, providing better feature initializations for our specific biological imaging context. For each MPI image, we extracted the features from the Hiera encoder, resulting in a feature tensor of dimension $256 \times 64 \times 64$. In order to analyze higher level features, we applied a 4×4 maxpool operation on these features, yielding reduced $256 \times 16 \times 16$ tensors. These were then flattened into 65536 dimension vectors. The feature vectors served as learned representations of the glomeruli activation patterns and were subsequently used in our alignment process. The extracted representations can formally be expressed as: $\mathbf{z}_t = \text{Flatten}(\text{Maxpool}(\mathbf{F}_t)) \in \mathbb{R}^d$, where $d = 65536$ is the dimension of our final feature representation.

3.3 AUTOENCODER

An autoencoder is a neural network that learns a compressed representation of input data by mapping it into a low-dimensional latent space and reconstructing the input from this space (Heaton, 2017). The architecture consists of two primary components: the encoder E and the decoder D.

The encoder maps the input image $\mathbf{x} \in \mathbb{R}^{H \times W \times C}$ to a latent representation $\mathbf{z} \in \mathbb{R}^d$, where $d \ll H \cdot W \cdot C$. This process is defined as $\mathbf{z} = E(\mathbf{x}; \theta_E)$, where θ_E represents the trainable parameters of the encoder. The decoder reconstructs the input from the latent representation: $\hat{\mathbf{x}} = D(\mathbf{z}; \theta_D)$, where $\hat{\mathbf{x}}$ is the reconstructed image, and θ_D denotes the decoder parameters. The autoencoder is trained to minimize the reconstruction loss, measured using the Mean Squared Error (MSE):

$$\mathcal{L}_{\text{recon}} = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2,$$

where N is the number of samples. By encoding information through latent space $\mathbf{z}$, the network learns a compact representation of input data.

3.4 OPTIMAL TRANSPORT

To align glomeruli activation patterns across two mice, we utilized the optimal transport (OT) framework with the Sinkhorn algorithm, which is a computationally efficient method for solving OT problems (Pham et al., 2020). The goal was to find the optimal mapping between two distributions while minimizing the transport cost. Fig. 3 (Appendix) depicts the joint pixel intensity distribution of corresponding glomeruli images for *33 dimethyl butyric acid* in the reference and target mice. The distribution highlights regions of high correlation (dark blue areas), suggesting consistent glomeruli activation patterns across the two mice. The concentration of the pixel intensities along the diagonal axis further indicates a strong alignment between the two sets of activation images. Outliers and lower-density regions (lighter areas) reveal minor mismatches, likely attributable to biological variability or noise in the data.

The OT approach can formally be expressed in the following manner. Given two sets of latent vectors, $\{\mathbf{z}_i^{(1)}\}$ for mouse 1 and $\{\mathbf{z}_j^{(2)}\}$ for mouse 2, and their corresponding probability distributions $\mathbf{p}$ and $\mathbf{q}$, the transport plan $\mathbf{T}$ is computed by solving:

$$\min_{\mathbf{T}} \sum_{i,j} \mathbf{T}_{ij} c(\mathbf{z}_i^{(1)}, \mathbf{z}_j^{(2)}) + \epsilon \sum_{i,j} \mathbf{T}_{ij} \log(\mathbf{T}_{ij}),$$

subject to the constraints $\mathbf{T}\mathbf{1} = \mathbf{p}, \quad \mathbf{T}^\top \mathbf{1} = \mathbf{q}$, where $c(\mathbf{z}_i^{(1)}, \mathbf{z}_j^{(2)})$ is the cost function (e.g., Euclidean distance) between latent vectors, and $\epsilon > 0$ is the entropy regularization parameter. The entropy regularization term, $\epsilon \sum_{i,j} \mathbf{T}_{ij} \log(\mathbf{T}_{ij})$, ensures numerical stability and faster convergence.

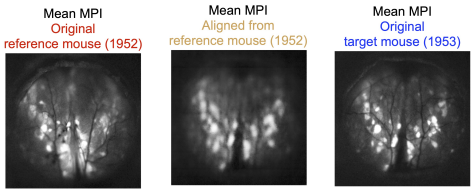

Figure 2: Mean MPI images reconstructed for the target mouse from aligned latent features of glomeruli activation patterns between reference and target mice.

By iteratively refining the transport plan $\mathbf{T}$, the Sinkhorn algorithm efficiently aligns the distributions of latent representations extracted from glomeruli activation images, revealing biologically meaningful correspondences. Its computational efficiency and robustness made it a key component of our alignment pipeline.

3.5 Approach and Evaluation Metrics

To quantitatively assess the success of our alignment process, we reconstructed the MPI images for the target mouse using aligned latent vectors obtained from the OT Sinkhorn framework. OT alignment was performed in the latent space of the autoencoder, where the Sinkhorn algorithm computed a transport plan that aligns the latent features of glomeruli activation patterns between the reference and target mice. Quantitative assessment of the alignment process was not performed for NNMF because we did not perceive any substantial changes between pre- and post-alignment (seen in Figure 4 [Appendix]), or for SAM2 because it lacked an out-of-box method to reconstruct encoded images.

Our approach can be summarized as follows. Let $\mathbf{Z}_{\text{ref}} \in \mathbb{R}^{n_{\text{ref}} \times d}$ represent the latent vectors extracted from the reference mouse, where n_{ref} is the number of samples, and d is the dimensionality of the latent space. The OT Sinkhorn algorithm computes a regularized transport plan $\mathbf{T} \in \mathbb{R}^{n_{\text{target}} \times n_{\text{ref}}}$ that minimizes the transport cost while satisfying marginal constraints. The aligned latent vectors $\mathbf{Z}_{\text{aligned}}$ for the target mouse are given as: $\mathbf{Z}_{\text{aligned}} = \mathbf{T} \cdot \mathbf{Z}_{\text{ref}}$, where $\mathbf{T}$ acts as the mapping between the reference and target latent spaces.

Once the aligned latent vectors $\mathbf{Z}_{\text{aligned}}$ are computed, they are passed through the decoder D of the autoencoder to reconstruct the corresponding MPI images. The decoder maps the aligned latent representations back to the pixel space: $\hat{\mathbf{X}}_{\text{recon}} = D(\mathbf{Z}_{\text{aligned}}; \theta_D)$, where $\hat{\mathbf{X}}_{\text{recon}} \in \mathbb{R}^{H \times W}$ is the reconstructed MPI image, and θ_D are the decoder parameters.

For each odorant, we reconstructed the images across $T = 10$ trials to account for variability. The final reconstructed MPI image is computed as the average over all trials for each odorant and mouse:

$$\bar{\mathbf{X}}_{\text{recon}} = \frac{1}{T} \sum_{t=1}^{T} \hat{\mathbf{X}}_{\text{recon},t},$$

where $\hat{\mathbf{X}}_{\text{recon},t}$ is the mean reconstructed image (mean MPI) for each odorant, as shown by the example in Fig. 2.

Subsequently, to compare the reconstructed mean MPI images $\bar{\mathbf{X}}_{\text{recon}}$ to the original mean MPI images of the reference $\bar{\mathbf{X}}_{\text{ref}}$ and target $\bar{\mathbf{X}}_{\text{target}}$ mice , we used the following metrics:

1) **Structural Similarity Index Measure (SSIM):** SSIM evaluates the perceptual similarity between two images $\mathbf{X}_1$ and $\mathbf{X}_2$ as: $\text{SSIM}(\mathbf{X}_1, \mathbf{X}_2) = \frac{(2\mu_1\mu_2 + C_1)(2\sigma_{12} + C_2)}{(\mu_1^2 + \mu_2^2 + C_1)(\sigma_1^2 + \sigma_2^2 + C_2)}$, where μ_1, μ_2 are means, σ_1^2, σ_2^2 are variances, σ_{12} is the covariance, and C_1, C_2 are small constants (Sara et al., 2019). The higher the value, the better the alignment between glomeruli activation pattern representations.

2) **Root Mean Squared Error (RMSE):** RMSE quantifies the pixel-wise differences between two images as: $\text{RMSE}(\mathbf{X}_1, \mathbf{X}_2) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\mathbf{X}_1[i] - \mathbf{X}_2[i])^2}$, where N is the total number of pixels

(Sara et al., 2019). The lower the value, the better the alignment between glomeruli activation pattern representations.

3) **Dice Coefficient:** The Dice coefficient measures the spatial overlap between two binary images A and B as: $\text{Dice}(A, B) = \frac{2|A \cap B|}{|A|+|B|}$, where $|A|$ and $|B|$ are the number of pixels in the binary images, and $|A \cap B|$ is the intersection of the two (Setiawan, 2020). The higher the value, the better the alignment between glomeruli activation pattern representations.

4 EXPERIMENTS AND RESULTS

To evaluate the efficacy of our OT-based alignment method, we used t-SNE plots to visualize representations of glomeruli activation patterns from reference and target mice. We compared t-SNE plots before and after applying the Sinkhorn algorithm.

For NNMF, MPI images were decomposed into their constituent features (basis components), which were aligned between the reference and target mice. The results are shown in Fig. 4 (Appendix). As demonstrated by the pre-alignment t-SNE plot, representations of the glomeruli activation patterns from different odorants and mice are scattered without any evident order. Following OT-based alignment using the Sinkhorn algorithm, the post-alignment t-SNE plot shows representations of glomeruli activation signals organized into tighter linear groups, but the improvement is not substantial. This is likely due to the fact that representations extracted by NNMF lacked the complexity required to capture subtle variations in the activation patterns for the two mice.

The autoencoder alleviated NNMF challenges, using LeakyReLU, Batch Normalization, and Dropout to minimize reconstruction loss and improve representation robustness. Image representations were extracted from the latent space. The pre-alignment t-SNE plot in Fig. 5 (Appendix) shows same-odorant glomeruli activations in separate clusters, with a separation between reference and target mouse signals, highlighting biological and experimental variability. After OT-based Sinkhorn alignment, the post-alignment t-SNE plot shows well-aligned same-odorant glomeruli activations forming tight clusters. The aligned data maintain their structure while reducing inter-mouse variability, with each odorant's activations mapped closer together. This demonstrates our alignment pipeline's effectiveness in managing biological variability and identifying consistent cross-subject activation patterns. The aligned images along with the original images for the reference and target mice are shown in Fig. 7 (Appendix) and Fig. 8 (Appendix). Quantitative assessment in Table 1 (Appendix) supports these findings, showing SSIM, RMSE, and DICE coefficient metrics for the aligned representations of 11 odorants. These were calculated using the images reconstructed from the aligned representation in relation to the average image for these odorants in the target mouse. All of the metrics showed improved alignment between reference and target mouse mean MPIs for all odorants.

SAM2 with its Hiera encoder offered an alternative approach for MPI image representation. The encoder's hierarchical structure effectively captured multi-scale glomeruli activation features. Qualitatively, the Hiera encoder demonstrated an ability to extract representations with a richer and more pronounced symmetric structure, as visualized in the t-SNE plot, compared to the NNMF and autoencoder approaches. This can be seen in Fig. 6 (Appendix). Similar to the autoencoder results in Fig. 5 (Appendix), the representations display an imaginary line dividing the mice, creating a symmetry axis for odorant clusters. Post-alignment results parallel the autoencoder's, indicating both methods facilitate effective OT.

5 DISCUSSION AND CONCLUSIONS

In this pilot, we aligned odorant glomeruli activation patterns between a reference and a target mouse. We extracted MPI image representations with NNMF, autoencoder, and SAM2. We spatially aligned them using the Sinkhorn OT algorithm. Both the t-SNE plots and quantitative metrics demonstrated successful alignment between reference and target animals, particularly with application of the autoencoder and SAM2. NNMF was a base case and was too simple to capture the complexity of the signal.

There are several limitations to our approach. In this paper we have considered only two animals, focused on a subset of odorants, and did not consider the temporal component in the alignment. A future extension of this study would involve expanding the number of animals and odorants, performing alignment with more complex odors composed of several odorants, spatially aligning only the regions of interest (ROIs) between animals rather than the entire image (useful for determining ligand protein interaction for olfactory receptors), and adding the temporal component in the alignment. Once alignment has been optimized, we would map odoroant glomeruli activation patterns from a reference mouse to multiple target mice and evaluate the accuracy of the mapping.

REFERENCES

Vivek Agarwal, Joshua Harvey, Dmitry Rinberg, and Vasant Dhar. Data science in olfaction, 2024. URL https://arxiv.org/abs/2404.05501.

LR Bijland, MK Bomers, and YM Smulders. Smelling the diagnosis: a review on the use of scent in diagnosing disease. *Netherlands Journal of Medicine*, 71:300–307, 2013.

Shawn D. Burton, Audrey Brown, Thomas P. Eiting, Isaac A. Youngstrom, Thomas C. Rust, Michael Schmuker, and Matt Wachowiak. Mapping odorant sensitivities reveals a sparse but structured representation of olfactory chemical space by sensory input to the mouse olfactory bulb. *eLife*, 11:e80470, 2022. doi: 10.7554/eLife.80470. URL https://doi.org/10.7554/eLife.80470.

Jerry D. Gibson and Alan Conrad Bovik. Handbook of image and video processing. 2000. URL https://api.semanticscholar.org/CorpusID:60713919.

Jay A. Gottfried. Central mechanisms of odour object perception. *Nature reviews. Neuroscience*, 11 (9):628–641, 2010. doi: https://doi.org/10.1038/nrn2883.

Jeff Heaton. Ian goodfellow, yoshua bengio, and aaron courville: Deep learning. *Genetic Programming and Evolvable Machines*, 19:305–307, 2017. URL https://api.semanticscholar.org/CorpusID:4300434.

D. Lee and H.S. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 399(6738):866–869, 1999.

Khiem Pham, Khang Le, Nhat Ho, Tung Pham, and Hung Hai Bui. On unbalanced optimal transport: An analysis of sinkhorn algorithm. In *International Conference on Machine Learning*, 2020. URL https://api.semanticscholar.org/CorpusID:211068892.

Baptiste Piqueret, Jean-Christophe Sandoz, and Patrizia d'Ettorre. The neglected potential of invertebrates in detecting disease via olfaction. *Frontiers in Ecology and Evolution*, 10:960757, 2023.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Dollár, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714*, 2024. URL https://arxiv.org/abs/2408.00714.

Chaitanya Ryali, Yuan-Ting Hu, Daniel Bolya, Chen Wei, Haoqi Fan, Po-Yao Huang, Vaibhav Aggarwal, Arkabandhu Chowdhury, Omid Poursaeed, Judy Hoffman, Jitendra Malik, Yanghao Li, and Christoph Feichtenhofer. Hiera: A hierarchical vision transformer without the bells-and-whistles. *ICML*, 2023.

Umme Sara, Morium Akter, and Mohammad Shorif Uddin. Image quality assessment through fsim, ssim, mse and psnr—a comparative study. *Journal of Computer and Communications*, 2019. URL https://api.semanticscholar.org/CorpusID:104425037.

T Sato, Y Katsuoka, K Yoneda, M Nonomura, S Uchimoto, R Kobayakawa, K Kobayakawa, and Y Mizutani. Sniffer mice discriminate urine odours of patients with bladder cancer: A proof-of-principle study for non-invasive diagnosis of cancer-induced odours. *Scientific Reports*, 7:14628, 2017.

T Sato, M Matsukawa, T Iijima, and Y Mizutani. Hierarchical elemental odor coding for fine discrimination between enantiomer odors or cancer-characteristic odors. *Frontiers Behavioral Neuroscience*, 16:849864, 2022.

Agung Wahyu Setiawan. Image segmentation metrics in skin lesion: Accuracy, sensitivity, specificity, dice coefficient, jaccard index, and matthews correlation coefficient. *2020 International Conference on Computer Engineering, Network, and Intelligent Multimedia (CENIM)*, pp. 97–102, 2020. URL https://api.semanticscholar.org/CorpusID:229703925.

ER Soucy, DF Albeanu, AL Fantana, VN Murthy, and Meister M. Precision and diversity in an odor map on the olfactory bulb. *Nature Neuroscience*, 12:210–220, 2009.

Joachim Weickert. Anisotropic diffusion in image processing. 1996. URL https://api.semanticscholar.org/CorpusID:47263773.

Im J. M. Kim M. Kim Y. R. Chung E. Yeon, C. Cranial and spinal window preparation for in vivo optical neuroimaging in rodents and related experimental techniques. *Experimental neurobiology*, 31(3):131–146, 2022. doi: https://doi.org/10.5607/en22015.

Jiayuan Zhu, Abdullah Hamdi, Yunli Qi, Yueming Jin, and Junde Wu. Medical sam 2: Segment medical images as video via segment anything model 2, 2024.

A APPENDIX

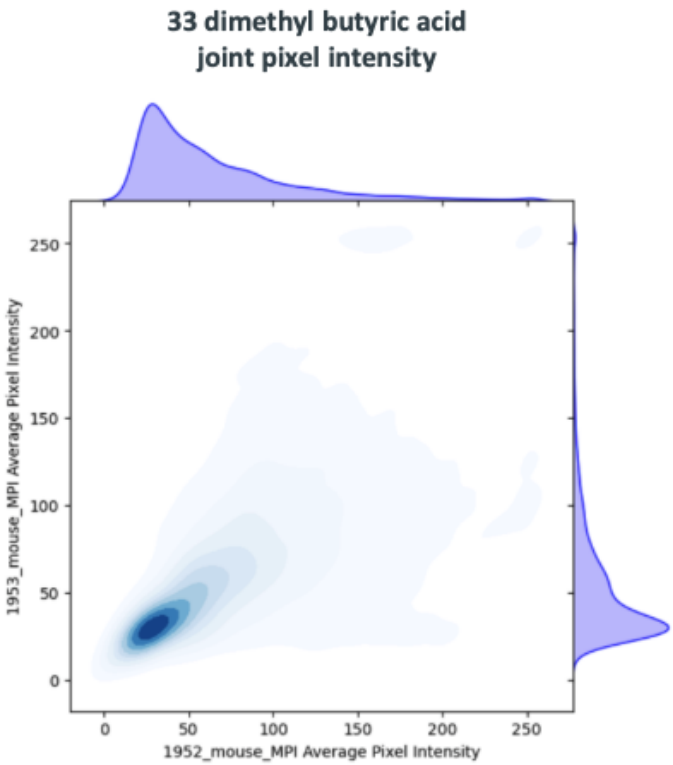

Figure 3: Joint pixel intensity distribution of glomeruli activation patterns for 33 dimethyl butyric acid.

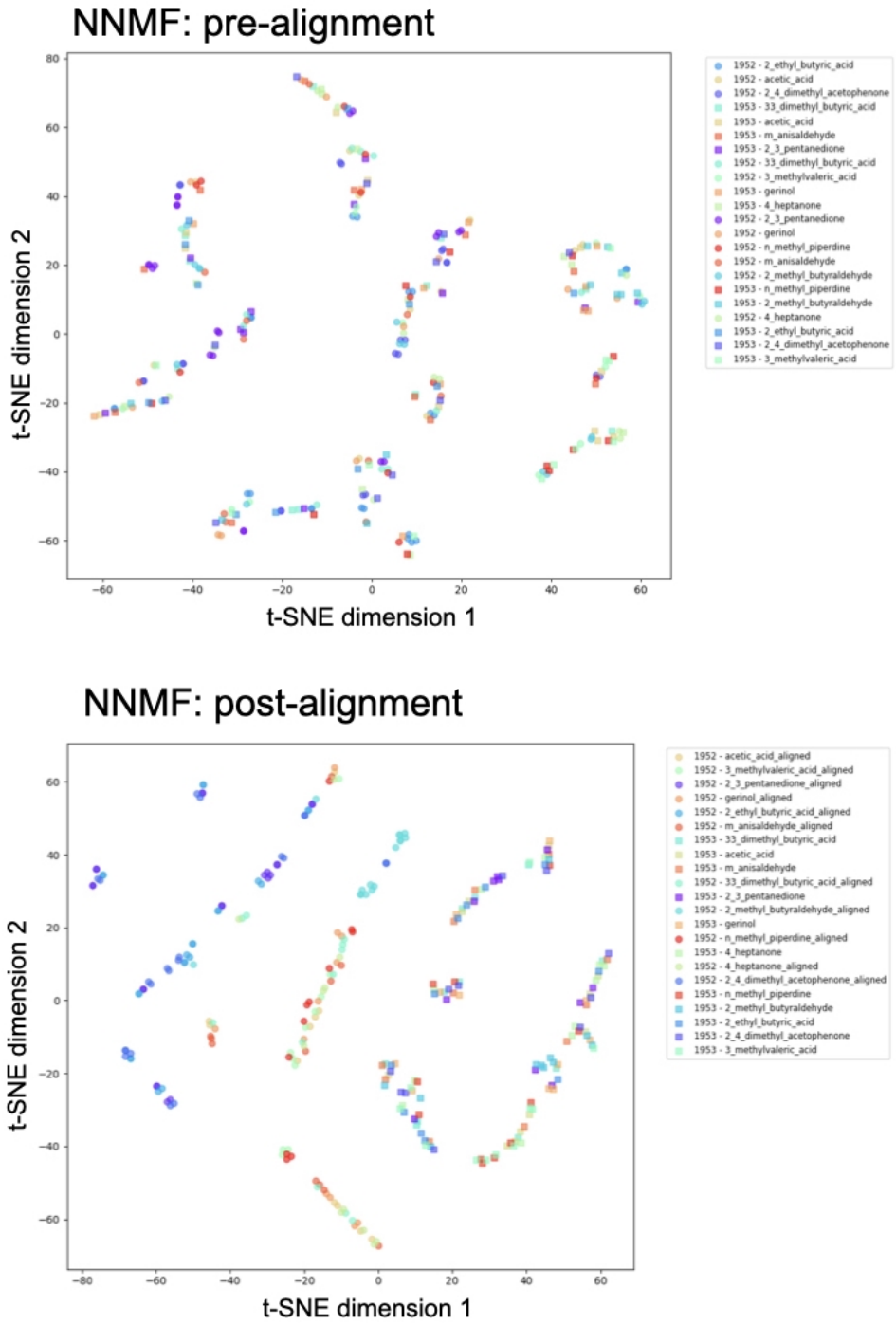

Figure 4: t-SNE visualization of NNMF representations before and after alignment. Each point represents the neural response pattern for a single odorant trial, color-coded by odorant identity. **Top panel (pre-alignment)**: Activation patterns corresponding to identical odorants across the two mice are scattered without forming distinct or coherent clusters, demonstrating that NNMF does not clearly separate odorant-specific neural responses across subjects. **Bottom panel (post-alignment)**: Even after alignment, activation patterns remain scattered and do not form clear clusters or overlapping structures. The NNMF-based alignment fails to effectively reduce inter-subject variability or clarify odorant-specific neuronal patterns.

The limited effectiveness of the NNMF-based alignment is due to its linear and shallow structure, restricting its ability to capture complex nonlinear relationships present in neural data. It does not leverage hierarchical or nonlinear transformations, making it less capable of extracting robust, generalizable latent representations for alignment across mice.

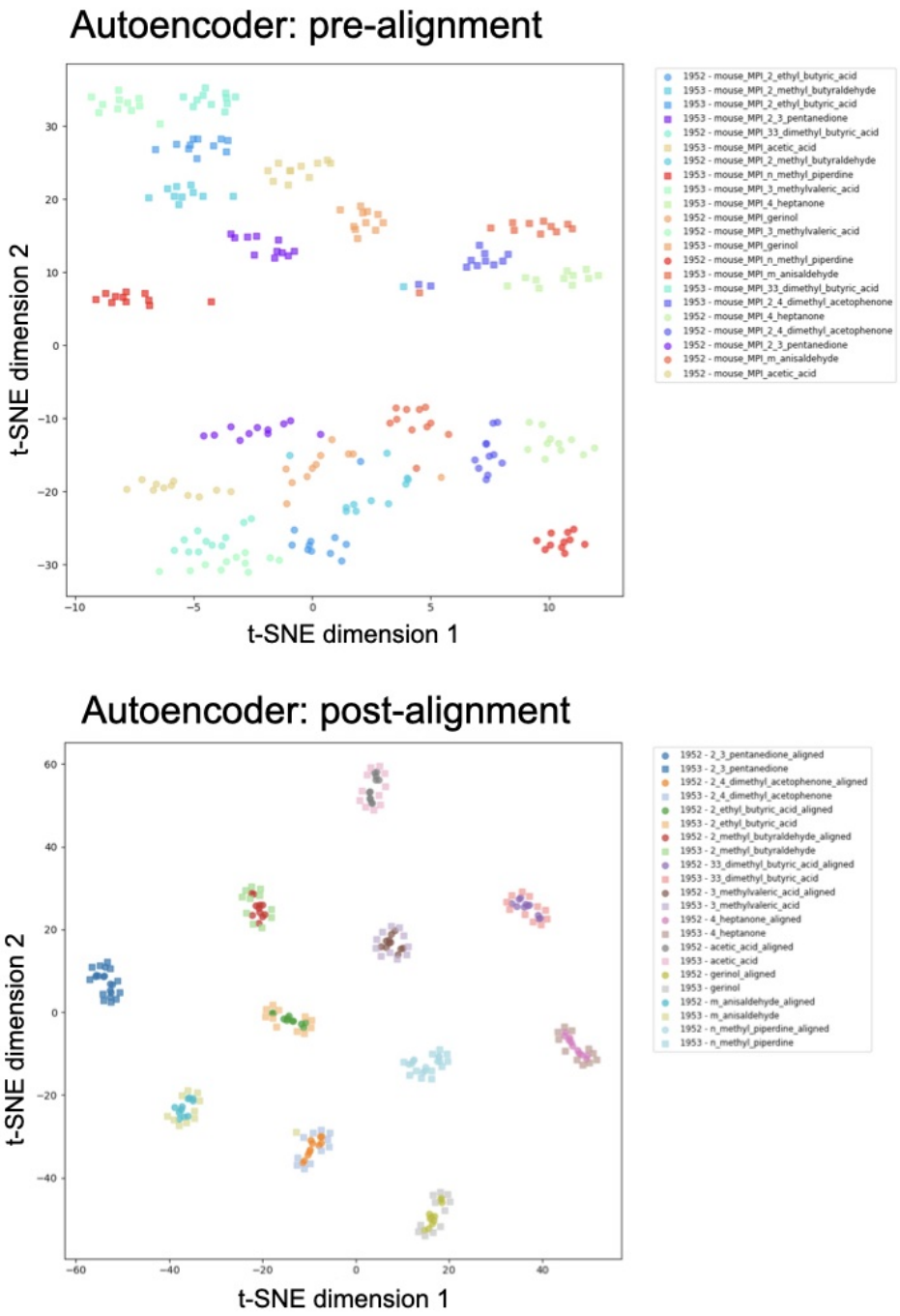

Figure 5: t-SNE visualization of autoencoder representations before and after alignment. Each point represents the neural activation pattern from a single odorant trial, color-coded by odorant identity. **Top panel (pre-alignment)**: Activation patterns corresponding to identical odorants from the two mice form separate, distinct clusters, reflecting significant inter-subject variability. **Bottom panel (post-alignment)**: After alignment, activation patterns from identical odorants across the two mice clearly overlap and form cohesive, well-defined clusters. The autoencoder-based alignment successfully reduces inter-subject variability, resulting in consistent, shared neuronal representations of odorants across animals.

The effectiveness of the autoencoder method arises from its deep, nonlinear architecture, enabling it to capture the complex hierarchical structures and nonlinear relationships present in neuronal activation data. Consequently, it generates robust and generalizable latent representations that facilitate successful alignment across individuals.

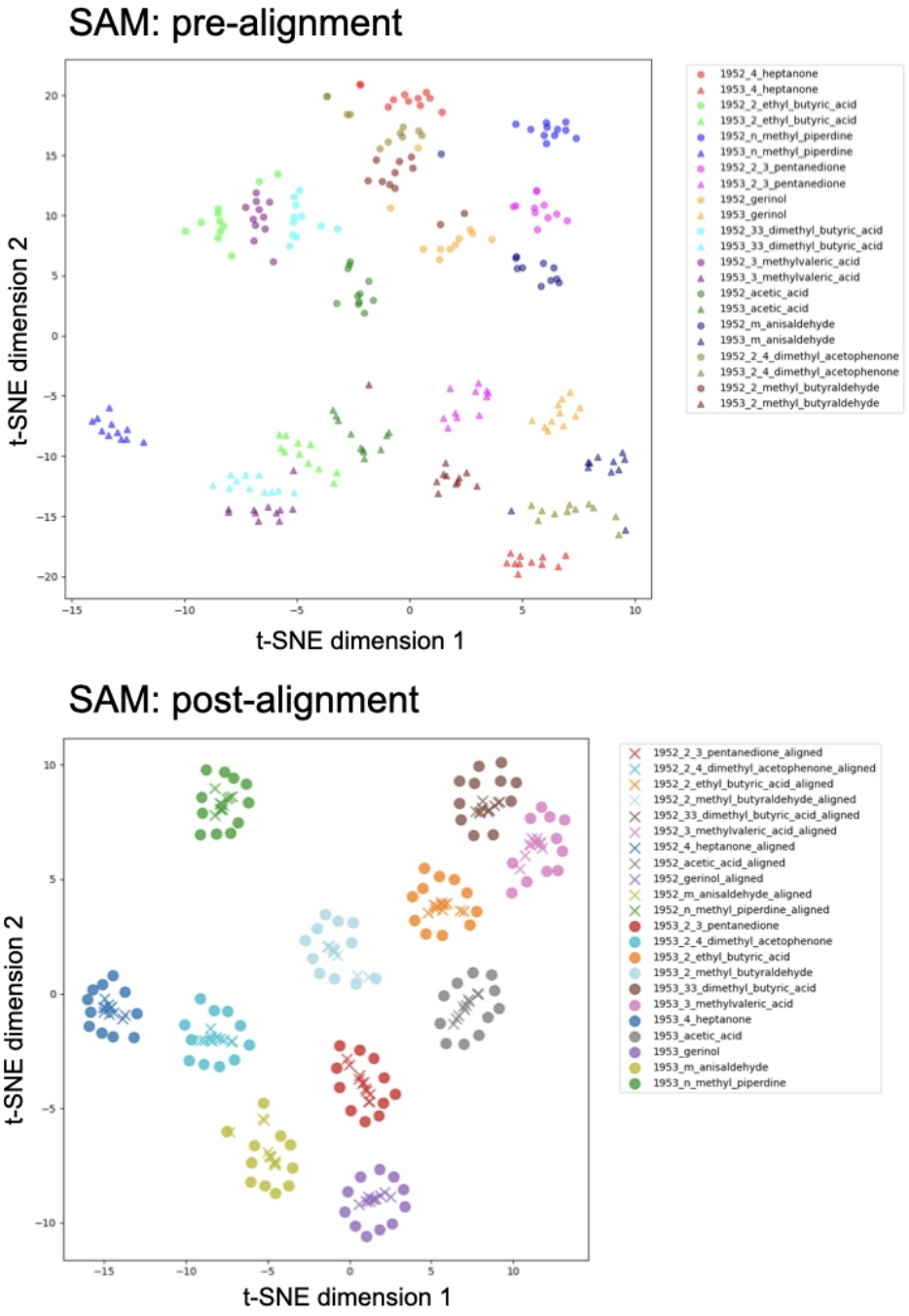

Figure 6: t-SNE visualization of SAM2 Hiera encoder representations before and after alignment. Each point corresponds to the neuronal activation pattern for a single odorant trial, color-coded by odorant identity.
The SAM-based alignment yields better clustering of odorant-specific neuronal representations. However, since it does not support reconstruction of the original neuronal activation images, we did not use it for further metrics based evaluation.

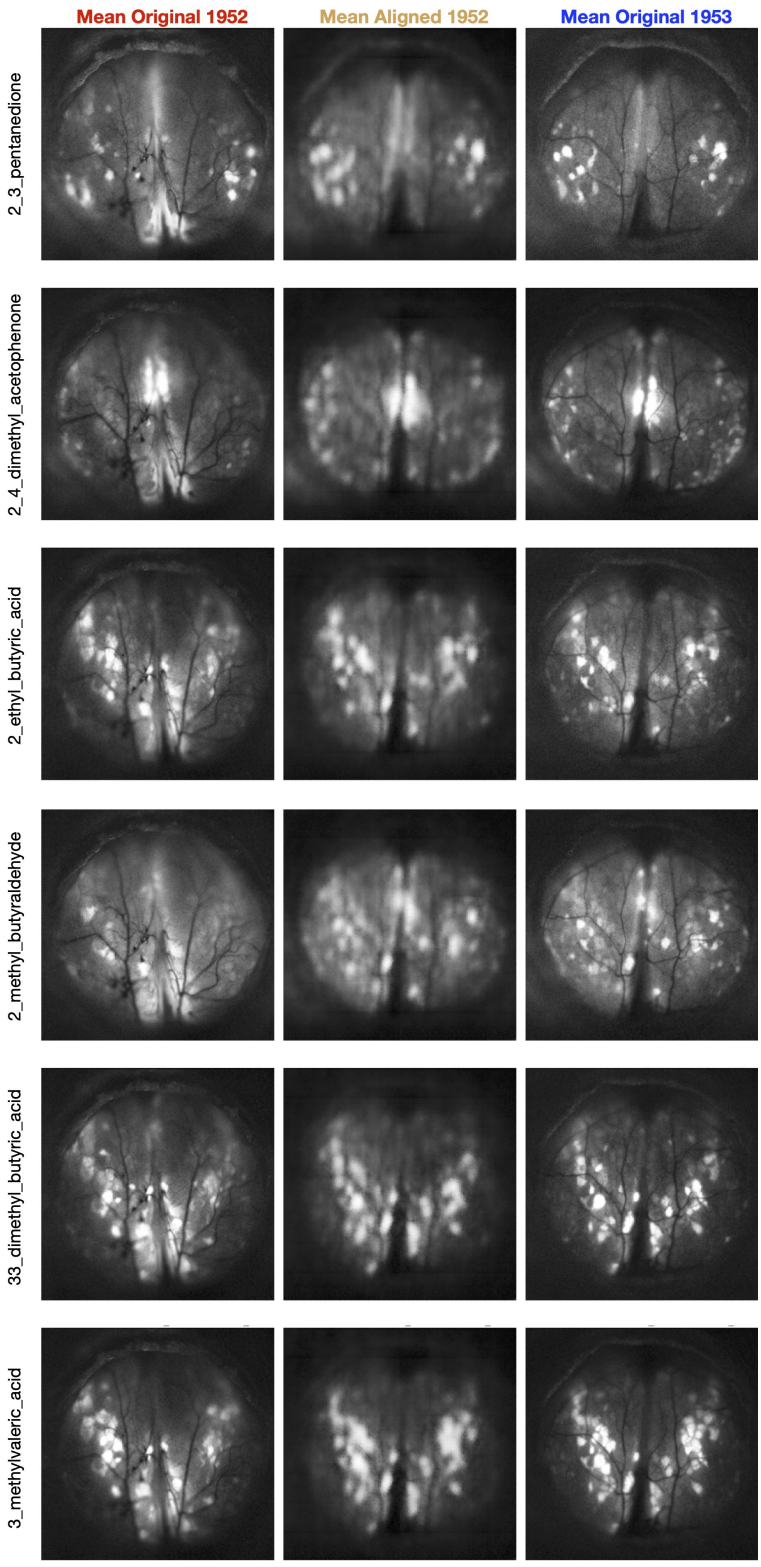

Figure 7: Aligned and original images for reference (1952) and target (1953) mice, continued.

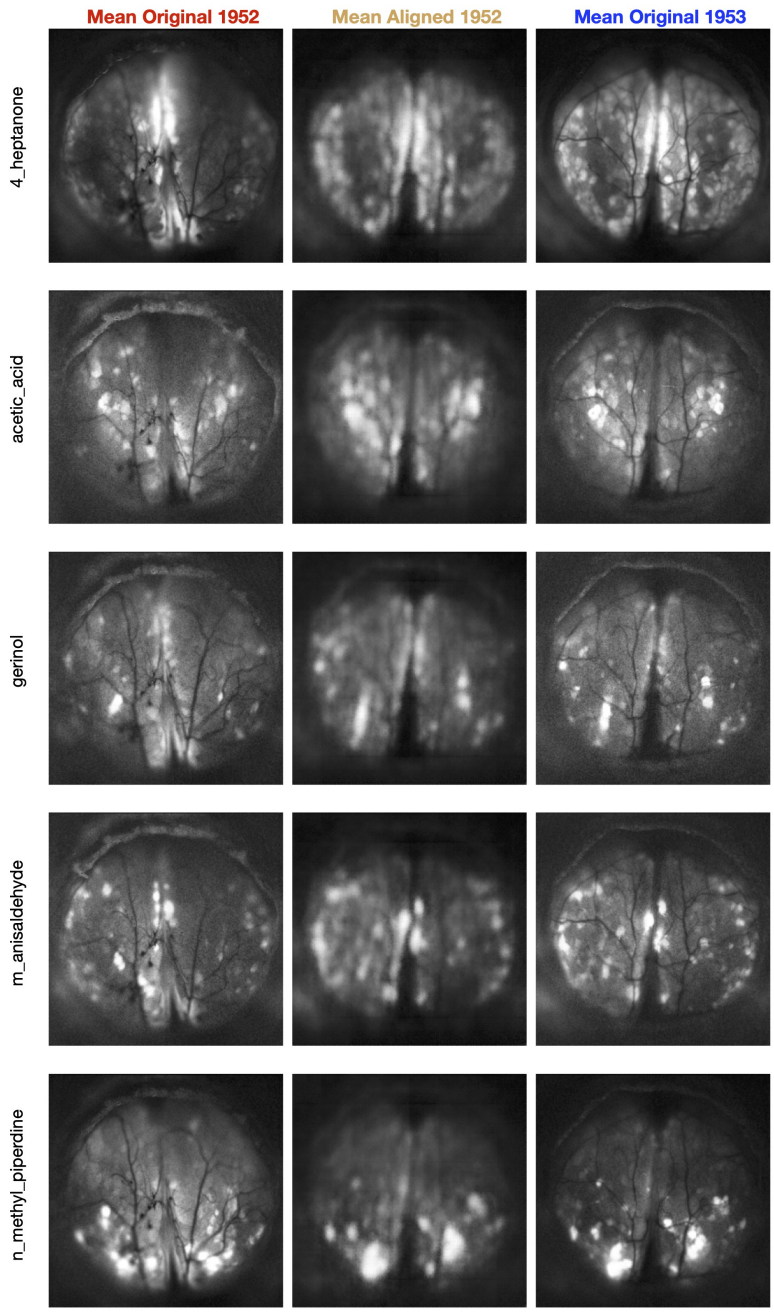

Figure 8: Aligned and original images for reference (1952) and target (1953) mice.

Odorant	Metric	Reference vs Target	Target vs Aligned
Autoencoder			
2_3_pentanedione	SSIM	0.2726	0.5089
	RMSE	0.1513	0.0568
	Dice	0.3366	0.3558
2_4_dimethyl_acetophenone	SSIM	0.3864	0.6749
	RMSE	0.1391	0.0499
	Dice	0.3046	0.3459
2_ethyl_butyric_acid	SSIM	0.3529	0.6079
	RMSE	0.1326	0.0436
	Dice	0.3501	0.3713
2_methyl_butyraldehyde	SSIM	0.3525	0.6338
	RMSE	0.1473	0.0465
	Dice	0.3724	0.3908
33_dimethyl_butyric_acid	SSIM	0.3891	0.6607
	RMSE	0.1428	0.0648
	Dice	0.3208	0.3442
3_methylvaleric_acid	SSIM	0.4447	0.6659
	RMSE	0.1371	0.0793
	Dice	0.3169	0.3493
4_heptanone	SSIM	0.3857	0.7077
	RMSE	0.1807	0.0523
	Dice	0.3388	0.4058
acetic_acid	SSIM	0.2679	0.5802
	RMSE	0.1279	0.0497
	Dice	0.3792	0.3849
gerinol	SSIM	0.2118	0.4526
	RMSE	0.1403	0.0821
	Dice	0.3645	0.3674
m_anisaldehyde	SSIM	0.2731	0.5871
	RMSE	0.1392	0.0471
	Dice	0.3363	0.3600
n_methyl_piperdine	SSIM	0.3795	0.6481
	RMSE	0.1694	0.0823
	Dice	0.3116	0.3178

Table 1: Metrics for autoencoder outputs of aligned representations in various odorants. The above table shows a comparison of metrics between the Reference vs Target mice and Target vs Aligned Glomeruli patterns of the reference mice. It shows an increase in values of SSIM, DICE scores and decrease in values of RMSE scores which indicates better aligned glomeruli patterns. (This analysis was done with the autoencoder because of its capability to reconstruct images given a latent space representation, and its latent space representation properties.)

For calculation of Structural Similarity Index Measure (SSIM), the terms μ and σ represent local statistical measures automatically computed within small windows across the image. μ refers to the local mean intensity within each window. It captures the local brightness similarity between the two images. σ refers to the local standard deviation and covariance within each window. It quantifies local contrast (variability in pixel intensity) and structural correlation between the corresponding regions of the two images.

The binarization threshold for DICE coefficient calculation was determined empirically by analyzing pixel intensity distributions and systematically evaluating thresholds from 0.1 to 0.9, both for background (vasculature) and the foreground (Glomerular activation).