A UNIFIED 2D-3D ENCODING AND JOINT GEOMETRIC METRIC FRAMEWORK FOR OLFACTORY MIXTURE PERCEPTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Existing olfactory perception methods often fail to capture critical three-dimensional spatial information and the distinction between causal odourants and environmental noise, limiting their ability to generalize across complex mixture scenarios. In this paper, we propose a unified molecular encoding framework that synergizes causal attention learning with a joint geometric metric space to accurately map molecular features to perceptual outcomes. First, we integrate 2D topology with 3D spatial configurations, constructing a model that effectively addresses the representation deficiencies inherent in traditional topological graphs. Second, we introduce a causal aggregation module reinforced by orthogonal constraints to explicitly disentangle true odourants from background solvents, ensuring robustness in data-scarce environments. Crucially, we propose a joint hyperbolic-cosine metric space for mixture similarity prediction; the hyperbolic component naturally encodes the hierarchical generative ancestry of odourants, whereas the cosine component captures the holistic, permutation-invariant nature of complex mixtures. Extensive experiments demonstrate that integrating these diverse geometric priors with causal representations yields state-of-the-art performance on complex mixture benchmarks while providing profound biological interpretability within the latent olfactory space. The code is available at `https://anonymous.4open.science/r/3D-2D-Geo-metric-mix-A2CE`.

1 INTRODUCTION

Smell plays a vital role in humans and animals Polster & Schieberle (2017); Dikeçligil & Gottfried (2024). It provides environmental warnings and regulates physiological states by affecting mood and appetite Chiera et al. (2025); Hao et al. (2025). A keen sense of smell is indispensable across various professions. Medically, olfactory sensing offers noninvasive early diagnosis via breath and skin volatiles, showing broad prospects in metabolic monitoring and detecting diseases such as lung cancer and schizophrenia D'Amico et al. (2008). Currently, the study of olfactory mechanisms is an important intersection of biology and artificial intelligence Hulsen et al. (2019); Dikeçligil & Gottfried (2024). Olfaction is initiated by odourants binding to receptors; however, deciphering the intricate connections between physicochemical properties and perceptual characteristics remains a fundamental challenge. Without a unified framework for olfactory coding Barwich & Lloyd (2022), accurate mapping of chemical structures to perceptual outcomes remains an elusive challenge in quantitative structure-odor relationship (QSOR) research Sagar et al. (2023); Trimmer et al. (2019); Sharma et al. (2021).

Early research primarily relied on hand-crafted expert descriptors in conjunction with classical machine learning algorithms—such as tree-based models, support vector machines, and linear models—to bridge chemical structural information with olfactory perception Vigneau et al. (2018); Snitz et al. (2013); Ravia et al. (2020). However, manual feature selection is heavily contingent upon prior domain knowledge. The mapping between the physicochemical odor space and perceptual descriptors is characterized by high-dimensional and nonlinear interactions. Traditional models, particularly linear and basic tree-based ones, struggle to simulate these complex, multilayered dependencies. This often leads to significant predictive conflicts when molecules that are structurally similar yet perceptually distinct are encountered. With the advancement of deep learning, data-driven methods have been

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

increasingly employed to decipher the mapping between the chemical structures/physicochemical properties of odourant molecules and their resulting olfactory perceptions. However, predicting odor profiles from molecular structures is fundamentally contingent upon the quality of molecular representations. Concurrently, the acquisition of olfactory data is both labor intensive and costly, necessitating rigorous training from panelists and the exclusion of toxic compounds. While the fragrance industry is reported to utilize between 10,000 and 20,000 distinct compounds, the largest publicly available dataset, GoodScents–Leffingwell (GS-LF), comprises only approximately 5,000 molecules. To facilitate the learning of robust molecular representations in these data-constrained scenarios, graph neural networks (GNNs) have been extensively adopted in the field of olfactory perception. The Principal Odor Map (POM) Lee et al. (2023) is a high-dimensional, data-driven representation of the olfactory perceptual space learned directly from molecular structure. While it demonstrates some performance in predicting the odor of individual molecules, it focuses primarily on 2D molecular structures. Compared with two-dimensional graphical representations, three-dimensional structures can more accurately depict the spatial arrangement of atoms and the geometric symmetry of molecules Huang et al. (2024). This higher accuracy leads to better generalization ability. Moreover, three-dimensional molecular design can identify various features, such as quantum properties, chirality, and binding affinity with specific proteins Zhang et al. (2023).

POMMIX Tom et al. (2025) is an extension of POM that leverages single-molecule olfactory data to construct representations for odourantant mixtures, enabling the modelling and analysis of mixture-level olfactory perception.

To address these challenges, we propose a framework that integrates 2D topology with 3D spatial configurations to construct high-fidelity molecular representations, effectively mitigating the expressivity deficits of purely topological methods. Building on this foundation, we introduce a causal attention learning (CAL) module for mixture aggregation. Unlike standard mechanisms vulnerable to statistical bias, CAL explicitly disentangles the representation into causal olfactory factors and environmental noise via orthogonality constraints. By filtering out spurious correlations, our model robustly captures complex nonlinear interactions within the latent space, accurately quantifying perceptual differences between mixtures without the need for explicit mixing rules. The main contributions of our study are as follows:

- We propose a unified molecular encoding framework that fuses 2D topology with 3D spatial conformations via group-symmetric SDEs. By explicitly capturing stereochemical properties and maintaining SE(3)-equivariance, this approach overcomes the expressivity limitations of purely topological graphs, providing a physically interpretable latent space that accurately reflects odourant–receptor binding mechanisms.
- We introduce a multihead causal attention mechanism with explicit orthogonality constraints to model the nonlinear nature of complex olfactory mixtures. This module robustly disentangled true causal perceptual drivers from environmental noise, effectively filtering out spurious correlations in data-scarce regimes without relying on explicit mixing rules.
- We formulate a joint hyperbolic cosine metric space for mixture similarity prediction. The hyperbolic component encodes the hierarchical, generative ancestry of odourants, whereas the cosine component captures the permutation-invariant, holistic nature of mixture perception. Integrating these geometric priors yields state-of-the-art performance on complex mixture benchmarks while providing profound biological interpretability.

2 Related Work

2.1 Olfactory Perception

In the field of olfactory perception, early methods relied on physicochemical features, classical machine learning algorithms, and similarity measures. Snitz *et al.* Snitz et al. (2013) proposed an angle-distance model based on the overall structural vector of odor mixtures. By optimizing 1,433 physicochemical descriptors and screening 21 core descriptors, they achieved effective prediction of the perceptual similarity of multicomponent odor mixtures. Keller *et al.* Keller et al. (2017), relying on the DREAM Olfactory Prediction Challenge, constructed a machine learning model based on 4,884 physicochemical features of 476 structurally and perception-diverse molecules and psychophysical data of 49 subjects, which can accurately predict odor intensity, pleasure, and 8 out

of 19 semantic descriptors. Ravia *et al.* Ravia et al. (2020) optimized a prediction model for odor structure and perception on the basis of 49,788 pairwise perceptual similarity scores of 199 subjects for 242 multicomponent odor molecules, integrating 21 physicochemical features into a single metric in radians to predict the perceptual similarity of multicomponent odor pairs. Although existing methods have achieved promising results, they still suffer from several inherent limitations. First, adjacency-matrix–based molecular representations are unable to encode atom-level and global 3D structural information, even though such atomic and 3D properties are key determinants of the binding affinity between odourant molecules and olfactory receptors Zhang et al. (2024b). Second, compared with the more challenging task of olfactory perception for molecular mixtures, most prior studies focus primarily on predicting the odor properties of individual molecules. Finally, current approaches do not sufficiently exploit the physicochemical properties of molecules, leading to incomplete and less expressive molecular representations.

Deep neural networks have been used to explore the relationship between molecular structure and odor mapping. Lee *et al.* Lee et al. (2023) constructed a POM through a graph neural network. This map not only unifies multiple tasks in odor perception but is also reliable for predicting the odor quality of unknown odor molecules. The POMMIX model developed by Tom *et al.* Tom et al. (2025) extended the application scope of POM from single molecules to mixture spaces by introducing a hierarchical inductive bias architecture, thereby enhancing the generalization performance of molecules across datasets. By systematically evaluating the alignment of molecular representations of pretrained transformers (MoLFormer) with human olfactory perception, Taleb *et al.* Taleb et al. (2024) demonstrated that models trained solely on chemical structure self-supervision can effectively predict expert-labelled odor labels, continuous perception scores, and odor similarity.

2.2 Deep learning-based molecular representation

Deep neural networks (DNNs), particularly graph neural networks (GNNs), have revolutionized olfactory modelling by directly encoding 2D molecular topologies. To optimize these representations, Shang *et al.* Shang et al. (2018) designed the EAGCN, substituting binary adjacency matrices with real-valued bond attention to enhance feature expressiveness. In data-scarce scenarios, self-supervised learning has proven highly effective: Taleb *et al.* Taleb et al. (2024) demonstrated that structurally trained transformers can robustly predict perceptual scores, whereas Wang *et al.* Wang et al. (2022) proposed MolCLR, leveraging graph augmentations and contrastive learning for generalizable property prediction. Bridging representations with biological mechanisms, Chithrananda *et al.* Chithrananda et al. (2024) improved interpretability by mapping molecules to auxiliary olfactory receptor activation profiles.

Despite these advances, 2D topological methods inherently neglect 3D spatial conformations (e.g., stereostructure and spatial charge), which are crucial biophysical determinants of olfactory receptor binding Liu et al. (2023); Chen et al. (2021). To address this, recent research has shifted toward structure-driven paradigms. Inspired by receptors acting as spatial filters, Shuvaev *et al.* Shuvaev et al. (2024) introduced DeepNose, which uses CNNs to extract 3D shape features. In 3D representation learning, Jiao *et al.* Jiao et al. (2023) utilized an equivariant energy-based model to capture geometric and physical properties. To fuse multimodal knowledge, Stark *et al.* Stärk et al. (2022) aligned 2D graphs and 3D structures via contrastive pretraining, and Zhu *et al.* Zhu et al. (2022) designed cross-modal generation and masked reconstruction tasks to learn complementary 2D–3D chemical features. Additionally, Lu *et al.* Lu et al. (2019) employed hierarchical networks directly on 3D conformations to model multilevel quantum interactions. Collectively, these works underscore the necessity of integrating multidimensional molecular configurations to accurately characterize olfactory mechanisms.

3 Problem Definition

The primary objective of olfactory perception modelling is to establish a computational framework that maps molecular structures to their perceptual representations, thereby enabling the precise quantification of odor similarities between complex mixtures. Let an odourant molecule be denoted as a structural graph $G = (\mathcal{V}, \mathcal{E}, R)$, encompassing both 2D topological connections and 3D spatial coordinates. We first define a monomolecular encoder Φ that maps G into a continuous high-dimensional embedding $z^{mono} = \Phi(G) \in \mathbb{R}^d$.

For a complex mixture $M = \{G_1, G_2, \ldots, G_n\}$, the human perceptual outcome is inherently nonlinear and nonadditive. It is typically dictated by a sparse subset of dominant causal odourants, while other components merely act as weak modifiers or environmental noise. Therefore, the core challenge lies in learning a causal aggregation function Ψ that processes the set of monomolecular embeddings to yield a robust mixture representation $z_{mix} = \Psi(\{z_1^{mono}, \ldots, z_n^{mono}\})$, which explicitly disentangles the invariant causal factors from suppressed background variations.

Ultimately, given two mixtures M_A and M_B, our goal is to estimate their perceptual similarity $S(M_A, M_B)$ on the basis of their aggregated representations z_{mix}^A and z_{mix}^B. To capture both the generative ancestry of odourants and the holistic nature of mixture perception, we formulate S within a joint geometric metric space integrating hyperbolic and cosine distances. The overall learning objective is to optimize the parameters of Φ and Ψ such that the predicted geometric distance tightly approximates the human-annotated perceptual similarity y_{sim}, formulated as minimizing the expected loss $\mathbb{E}_{\mathcal{D}}[\mathcal{L}(S(M_A, M_B), y_{sim})]$ subject to causal regularization constraints.

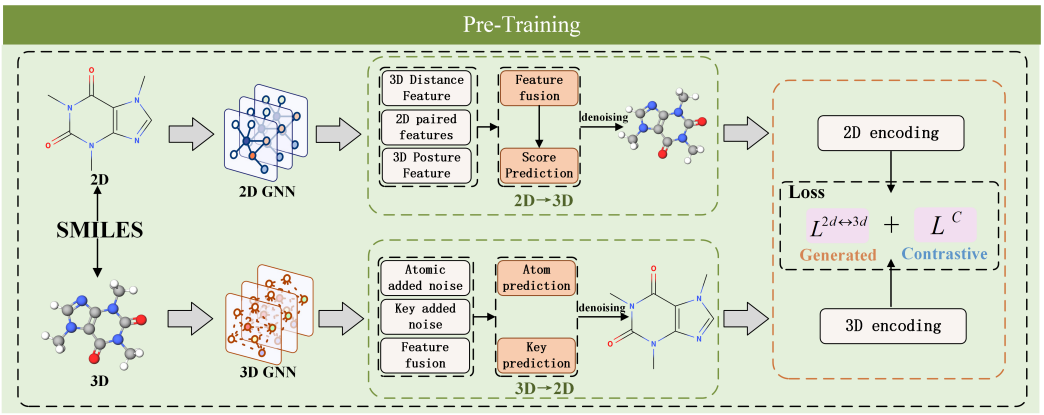


Figure 1: 2D-3D Information Fusion Framework.

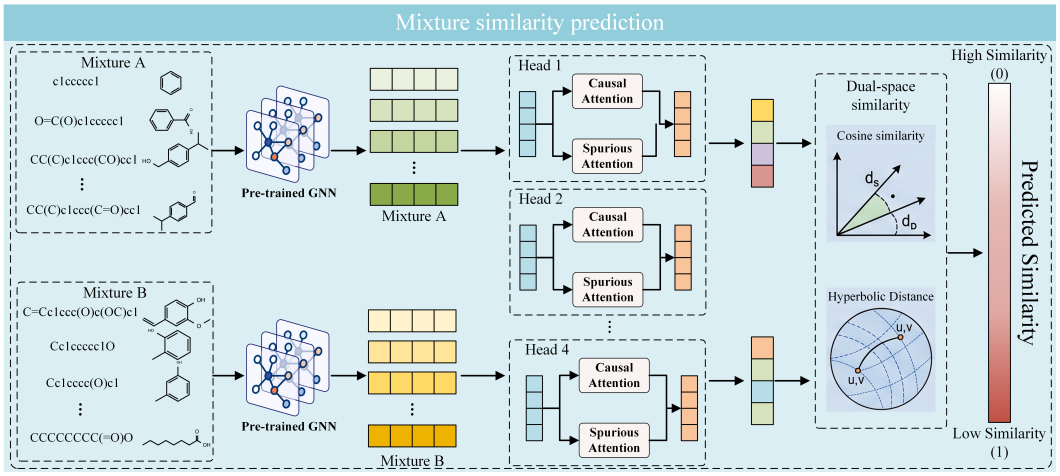


Figure 2: Framework for Mixture Similarity.

4 Framework

4.1 Group-Symmetric SDE for 2D-3D Fusion

To rigorously capture the interplay between chemical topology and spatial arrangement, we construct a continuous generative framework governed by group symmetric stochastic differential equations (SDEs). This framework ensures that the learned representations respect the fundamental physical symmetries of molecules, specifically SE(3)-equivariance and reflection-antisymmetry.

The spatial interactions between atoms are highly sensitive to interatomic distances. To capture these high-frequency geometric variations, we encode the Euclidean distance $d_{ij} = \|r_i - r_j\|_2$ between connected atoms i and j using a set of Gaussian radial basis functions. The distance embedding $\mathbf{e}_{dist}^{ij} \in \mathbb{R}^K$ is explicitly defined as follows:

$$\mathbf{e}_{dist,k}^{ij} = \exp\left(-\gamma_k (d_{ij} - \mu_k)^2\right), \quad k = 1, \ldots, K \tag{1}$$

where μ_k and γ_k are learnable centers and bandwidths, respectively. This continuous distance encoding is then fused with discrete topological edge features $\mathbf{e}_{topo}^{ij}$ via an elementwise gating mechanism to form the edge representation:

$$\mathbf{h}_{edge}^{ij} = \text{MLP}(\mathbf{e}_{topo}^{ij}) \odot \mathbf{e}_{dist}^{ij} + \mathbf{e}_{3D}^{ij} \tag{2}$$

This ensures that the model is sensitive to subtle conformational changes critical for olfactory receptor binding.

A core challenge in 3D molecular generation is ensuring that the generated structures are valid regardless of the global orientation. We resolve this by constructing equivariant local frames on each chemical bond, allowing the model to process geometric features in a local, invariant coordinate system. For a pair of atoms i and j with coordinates $r_i, r_j \in \mathbb{R}^3$, we define an orthonormal basis $\mathcal{F}_{ij} = (\vec{e}_1, \vec{e}_2, \vec{e}_3)$ centered at the bond. Following the Gram–Schmidt process adapted for molecular graphs [21], the frame basis vectors are constructed as follows:

$$\vec{e}_1 = \frac{r_i - r_j}{\|r_i - r_j\|}, \quad \vec{e}_2 = \frac{r_i \times r_j}{\|r_i \times r_j\|}, \quad \vec{e}_3 = \vec{e}_1 \times \vec{e}_2 \tag{3}$$

where $\times$ denotes the cross product. Crucially, $\vec{e}_2$ and $\vec{e}_3$ involve the cross product, making them pseudovectors that flip signs under reflection. This property enables the model to distinguish enantiomers, which is vital for olfactory perception modelling since stereoisomers frequently possess distinct odor profiles.

Utilizing these local frames, we project equivariant 3D vector features $\vec{v} \in \mathbb{R}^3$ into invariant scalars $f \in \mathbb{R}^3$ to be processed by standard neural networks:

$$f = \text{Project}(\vec{v}, \mathcal{F}_{ij}) = [\vec{v} \cdot \vec{e}_1, \vec{v} \cdot \vec{e}_2, \vec{v} \cdot \vec{e}_3]^\top \tag{4}$$

Conversely, the scalar scores predicted by the network are backprojected into the global 3D space to update the coordinates:

$$\vec{v}' = \text{BackProject}(f', \mathcal{F}_{ij}) = f_1' \vec{e}_1 + f_2' \vec{e}_2 + f_3' \vec{e}_3 \tag{5}$$

This frame-transition mechanism guarantees that the final score function $S_\theta(x_t, t)$ is strictly SE(3)-equivariant.

We model the generation of 3D coordinates R conditioned on the 2D graph G_{2D} as learning the conditional score $\nabla_{R_t} \log p_t(R_t|G_{2D})$. During training, we sample a time step $t \sim \mathcal{U}(0, T)$ and perturb the ground-truth coordinates R_0 via a forward diffusion process determined by the SDE, resulting in a noisy sample R_t drawn from the transition kernel $p_{0t}(R_t|R_0)$. The objective is to minimize the expected squared Euclidean distance between the output of our SE(3)-equivariant score network S_θ^{eq} and the analytical score of the perturbation kernel:

$$\mathcal{L}^{2D\to 3D} = \mathbb{E}_{t,R_0,R_t\sim p_{0t}(R_t|R_0)} \left[\lambda(t) \left\| S_\theta^{eq}(G_{2D}, R_t, t) - \nabla_{R_t} \log p_{0t}(R_t|R_0) \right\|_2^2\right] \tag{6}$$

where $\lambda(t)$ is a positive weighting function to balance the magnitude of loss across different noise scales.

Conversely, to reconstruct the 2D topology from the 3D conformation, we employ an SE(3)-invariant diffusion process on the continuous relaxation of the discrete graph features (node attributes X and bond attributes E). Let $\mathbf{x} = (X, E)$ denote the unified 2D feature set. The forward process diffuses $\mathbf{x}_0$ into $\mathbf{x}_t$ using independent Brownian motions. The reconstruction objective uses an SE(3)-invariant score network S_θ^{inv} to estimate the gradient field required to denoise the topology:

$$\mathcal{L}^{3D\to 2D} = \mathbb{E}_{t,\mathbf{x}_0,\mathbf{x}_t\sim p_{0t}(\mathbf{x}_t|\mathbf{x}_0)} \left[\lambda(t) \left\| S_\theta^{inv}(R, \mathbf{x}_t, t) - \nabla_{\mathbf{x}_t} \log p_{0t}(\mathbf{x}_t|\mathbf{x}_0) \right\|_2^2\right] \tag{7}$$

This objective explicitly forces the latent representation to retain sufficient information to recover the chemical identity solely from the geometric configuration.

4.2 Multihead Causal Attention for Mixture Similarity

Human olfactory perception of mixtures is inherently nonlinear and nonadditive: typically, only a sparse subset of odourants governs the perceptual outcome, while other constituents act as neutral solvents, weak modifiers, or masking agents. Therefore, mixture learning must satisfy two principles: causal selectivity toward dominant odourants and invariance to environmental noise. To achieve this goal, we replace standard pooling with a novel multihead causal attention mechanism coupled with orthogonality constraints.

Let $z_i^{mono} = \Phi(G_i) \in \mathbb{R}^d$ denote the monomolecular embedding obtained from the frozen 2D-3D encoder. To contextualize the features before mixture aggregation, we first apply a nonlinear projection with a residual connection:

$$h_i = z_i^{mono} + \mathrm{LayerNorm}(\mathrm{ReLU}(W_{ctx} z_i^{mono})) \tag{8}$$

To capture the diverse and multifaceted nature of olfactory profiles, we partition the contextualized embedding h_i into K independent subspace heads such that $h_{i,k} \in \mathbb{R}^{d/K}$. Within each head k, we introduce two parallel attention pathways: a causal branch designed to extract the dominant perceptual drivers and a suppressive branch intended to capture the background or environmental noise. The attention weights for molecule i in mixture M are computed as follows:

$$\alpha_{i,k}^c = \mathrm{Softmax}(\mathrm{MLP}_c(h_{i,k})), \quad \alpha_{i,k}^s = \mathrm{Softmax}(\mathrm{MLP}_s(h_{i,k})) \tag{9}$$

where MLP_c and MLP_s map the subspace representations to unnormalized scalar scores. The aggregated causal and suppressive representations for the mixture in head k are then obtained via weighted summation:

$$z_{mix,k}^c = \sum_{i \in M} \alpha_{i,k}^c h_{i,k}, \quad z_{mix,k}^s = \sum_{i \in M} \alpha_{i,k}^s h_{i,k} \tag{10}$$

To ensure strict disentanglement between the causal odor profile and the irrelevant background components, we impose a headwise orthogonality constraint. Specifically, we minimize the squared cosine similarity between the aggregated causal and suppressive representations:

$$\mathcal{L}_{orth} = \frac{1}{K} \sum_{k=1}^{K} \cos^2(z_{mix,k}^c, z_{mix,k}^s) \tag{11}$$

This regularization forces the network to dynamically separate invariant olfactory determinants from nuisance variations. Components strongly aligned with the dominant causal subspace accumulate constructively, whereas environmental factors are pushed into an orthogonal latent space.

Finally, the dominant causal features from all K heads are concatenated and linearly projected to form the robust, unified representation of the mixture:

$$z_{mix} = W_{proj}[z_{mix,1}^c \parallel z_{mix,2}^c \parallel \cdots \parallel z_{mix,K}^c] \tag{12}$$

The overall training objective integrates the task prediction error and the causal disentanglement penalty:

$$\mathcal{L} = \mathbb{E}_{M_A, M_B} \left[||S(M_A, M_B) - y_{sim}||_2^2 + \lambda \mathcal{L}_{orth} \right] \tag{13}$$

where $S(M_A, M_B)$ represents the final predicted similarity—computed as a learned weighted combination of the cosine similarity and hyperbolic distance between z_{mix}^A and z_{mix}^B (detailed in Section 4.3)—and λ controls the strength of the orthogonal constraint.

4.3 Olfactory Perceptual Similarity Prediction

Olfactory perception is not primarily organized by explicit molecular descriptors but instead emerges from the statistical structure of odor co-occurrence in natural environments. At the biological level, odourants are produced through metabolic pathways that form inherently hierarchical reaction networks. Starting from a limited set of precursor compounds, successive enzymatic transformations generate increasingly specialized molecules, giving rise to a branching structure that can be approximated by a tree $\mathcal{T}$. Each odourant corresponds to a leaf node, and their similarity is governed by shared ancestry within this hierarchy. This induces a natural tree metric

$$d_{\mathcal{T}}(i, j) = \mathrm{depth}(i) + \mathrm{depth}(j) - 2 \cdot \mathrm{depth}(\mathrm{LCA}(i, j)), \tag{14}$$

where $\mathrm{LCA}(i, j)$ denotes the lowest common ancestor. Under this formulation, odourants that share longer biochemical pathways exhibit greater similarity and are more likely to cooccur.

Such hierarchical generative processes impose strong geometric constraints. Tree structures exhibit exponential growth, $N(r) \sim b^r$, reflecting the combined expansion of metabolic branches, whereas Euclidean spaces grow polynomially, $V_{\mathbb{E}}(r) \sim r^d$, making distortion-free embedding fundamentally impossible. This mismatch is not merely quantitative but topological: Euclidean representations fail to preserve higher-order connectivity patterns observed in odor co-occurrence data, cwhich is consistent with empirical findings that they cannot reproduce the corresponding Betti curves. In contrast, hyperbolic space provides natural continuous embedding for hierarchical structures because of its exponential volume growth, $V_{\mathbb{H}}(r) \sim e^{(d-1)r}$. More formally, the geodesic distance in hyperbolic space is given by

$$\cosh d_{\mathbb{H}}(x, y) = \cosh r_x \cosh r_y - \sinh r_x \sinh r_y \cos(\Delta\theta), \tag{15}$$

which decomposes into a radial component (encoding hierarchical depth) and an angular component (capturing intralevel similarity), thereby closely approximating the tree metric. Geodesics in hyperbolic space naturally traverse toward a shared ancestor before diverging, making it a continuous analogue of metabolic hierarchies and explaining its ability to preserve both metric and topological properties of odor data.

While hyperbolic distance captures the hierarchical and generative structure of the odor space, human olfactory perception has an additional defining characteristic: it operates on odor mixtures as holistic perceptual objects rather than decomposing them into individual molecular components. Consequently, perceptual similarity depends on the overall compositional pattern of a mixture rather than the absolute magnitude of its constituents. To capture this property, we represent each odor mixture as a vector in a high-dimensional space and employ cosine similarity

$$\mathrm{sim}_{\cos}(z_A, z_B) = \frac{z_A \cdot z_B}{\|z_A\|\|z_B\|}, \tag{16}$$

which measures directional alignment while being invariant to the vector norm. This invariance is crucial in the olfactory setting, as the norm may reflect concentration, number of components, or intensity, whereas perception is governed primarily by relative compositional structure. By focusing solely on direction, cosine similarity effectively captures the global perceptual quality of mixtures independent of scale.

Moreover, cosine similarity naturally satisfies permutation invariance—a fundamental requirement for unordered odor mixtures. By evaluating the direction of an aggregated mixture vector, it captures the compositional invariance of olfactory processing, while its symmetry, $\mathrm{sim}(A, B) = \mathrm{sim}(B, A)$, aligns perfectly with human perceptual judgments. Crucially, cosine similarity mitigates the curse of dimensionality that plagues Euclidean metrics. By emphasizing relative feature overlap via normalized inner products, it ensures that perceptual distinctions remain highly discriminative within the sparse, high-dimensional latent spaces characteristic of complex mixtures.

5 Experiments

5.1 Experimental Setting

Our experiments are conducted on four representative olfactory datasets: GS-LF Flavor (2018), Snitz Snitz et al. (2013), Ravia Ravia et al. (2020), and BushdidBushdid et al. (2014). Among them, GS-LF is a curated combination of the GoodScents and Leffingwell databases and contains 4,983 odourant molecules with 138 expert-defined odour descriptors, providing a standard benchmark for molecular odour description prediction. Owing to the multilabel nature of odor annotation, this dataset exhibits substantial label imbalance and a pronounced long-tail distribution. The other three datasets, Snitz, Ravia, and Bushdid, focus on human olfactory perception rather than expert semantic annotation. Snitz provides averaged perceptual similarity ratings for 359 unique odourant pairs from 139 participants, whereas Ravia contains 195 unique pairs of multicomponent and monomolecular odour stimuli rated by 94 participants. Bushdid is constructed from a triangle test paradigm in which participants identify odd samples among three mixtures, thereby measuring perceptual discriminability between complex odor mixtures. Taken together, these four datasets allow us to assess the model from complementary perspectives, including expert-labelled odor descriptor prediction, perceptual

similarity modelling, and mixture discriminability evaluation. We evaluate our approach on the mixture dataset by training and testing on 5-fold cross-validation (CV) splits, stratified across the Snitz, Ravia, and Bushdid datasets. For early stopping, a validation split is randomly split from the training set, producing a final split of 70/10/20 training/validation/test sets. The performances of the models are then evaluated on the test sets. After pretraining our model, the 2D-GNN network fuses the 3D-2D information; then, we fine-tune it on the single-molecule data to fit the olfactory task, and finally, we freeze the GNN network for testing on the mixed-molecule dataset.

5.2 Experimental Results

Table 1: Performance of the pretrained model on the single-molecule dataset.

Model	**Test AUROC**
GCN	0.8508
GAT	0.8553
GIN	0.8558
MPNN	0.8575
MPNNPOM	0.8632
DenseSense	0.8530
Mol-PECO	0.7969
Ours	**0.8763**

Table 2: Model performances on mixture data. Fivefold cross-validation metrics are reported for all baseline models.

Model	Pearson ρ (↑)	RMSE (↓)	Kendall τ (↑)
Snitz baseline	0.399	0.334	0.292
XGBoost + MolT5	0.432	0.171	0.347
XGBoost + RDKIT	0.485	0.166	0.373
XGBoost + POM	0.497	0.165	0.388
POMMix	0.727	0.131	**0.542**
Ours	**0.779**	**0.131**	0.534

To comprehensively evaluate the effectiveness of our 3D-fused molecular representation, we benchmark it against a diverse suite of representative baselines. For single-molecule tasks, these include general-purpose graph neural networks (GCN, GAT, and GIN, implemented via PyTorch Geometric), message passing architectures (a generic MPNN and MPNNPOM), and specialized olfactory approaches (DenseSense Saha et al. (2025) and Mol-PECO Zhang et al. (2024a)). Furthermore, to assess our proposed mixture similarity prediction framework, we compare its performance against established mixture-level baselines, including XGBoost, Snitz, and POMMix.

As shown in Table 1, our model achieves a state-of-the-art test AUROC of 0.8763 on the single-molecule dataset. This superiority highlights the advantage of our group-symmetric SDE framework, which deeply fuses 2D topology with 3D spatial conformations. Unlike purely topological baselines, our model explicitly captures stereochemical properties and high-frequency interatomic distances, effectively overcoming their expressivity limitations and providing a physically interpretable latent space that better reflects odourant–receptor binding mechanisms.

For complex mixtures (Table 2), our framework yields the highest Pearson correlation of 0.779 and the lowest RMSE of 0.131, significantly outperforming sequence- or 2D-based baselines. This breakthrough is driven by two key innovations. First, the multihead causal attention mechanism explicitly disentangles dominant causal odourants from environmental noise via orthogonality constraints, ensuring robust aggregation in data-scarce regimes without relying on rigid mixing rules. Second, our joint hyperbolic cosine metric space accurately calibrates absolute perceptual distances. The hyperbolic component naturally encodes the hierarchical generative ancestry of odourants, whereas the cosine similarity encapsulates the holistic, permutation-invariant nature of mixture perception.

Consequently, our joint-metric approach—coupled with causal representations—substantially improves upon baselines such as POMMix, precisely mapping structural features to perceptual outcomes within a biologically plausible continuous manifold.

5.3 Ablation Study

Table 3: Ablation study results of different distance metrics across 5-fold cross-validation.

Fold	Hyperbolic distance			Cosine Similarity			Ours		
	PLCC	KLCC	RMSE	PLCC	KLCC	RMSE	PLCC	KLCC	RMSE
cv0	0.745	0.486	0.167	0.671	0.440	0.259	**0.787**	**0.511**	**0.134**
cv1	0.382	0.464	0.451	0.623	0.451	0.247	**0.780**	**0.537**	**0.124**
cv2	0.590	0.469	0.202	0.652	0.467	0.252	**0.749**	**0.525**	**0.142**
cv3	0.658	0.513	0.198	0.738	0.539	0.218	**0.826**	**0.589**	**0.125**
cv4	0.687	0.508	0.164	0.650	0.445	0.238	**0.753**	**0.508**	**0.127**
Average	0.612	0.488	0.236	0.667	0.468	0.243	**0.779**	**0.534**	**0.131**

To systematically validate the necessity of our proposed joint metric framework, we conducted a 5-fold cross-validation ablation study, as summarized in Table 3, and compared the predictive performance of purely hyperbolic distance, purely cosine similarity, and our combined approach. Empirical results indicate that relying on a single geometric metric yields suboptimal representations that fail to fully capture the complexity of olfactory perception. Specifically, using only the hyperbolic distance achieves an average PLCC of 0.612 and an RMSE of 0.236; however, while it naturally encodes the hierarchical generation of odourants and metabolic pathways, it lacks the capacity to describe holistic, permutation-invariant compositional patterns. Conversely, purely cosine similarity improves the average PLCC to 0.667 but yields a lower KLCC and higher RMSE, as it excels at measuring concentration-independent directional alignment but fundamentally ignores topological ancestry and hierarchical priors. By unifying these complementary metrics, our combined approach demonstrates a strong synergistic effect, reaching the highest average PLCC of 0.779 and KLCC of 0.534 while drastically reducing the average RMSE to 0.131. This substantial improvement confirms that integrating diverse geometric priors is essential for accurately calibrating human olfactory perceptual distances, as the hyperbolic component anchors latent representations within a biologically plausible hierarchical manifold, whereas the cosine component sharpens mixture-level perceptual boundaries on the basis of holistic compositional similarity.

6 Conclusion and Discussion

In this paper, we propose a unified molecular encoding framework that integrates 2D topology with 3D spatial conformations, combined with causal attention learning under orthogonality constraints, to precisely capture dominant odourants and effectively decouple environmental noise. Experimental results on multiple datasets—including GS-LF, Snitz, Ravia, and Bushdid—demonstrate that our approach achieves state-of-the-art performance in terms of perceptual prediction accuracy. The core innovation lies in the construction of a joint hyperbolic-cosine metric space, which leverages hyperbolic geometry to encode the hierarchical priors of metabolic pathways while utilizing cosine similarity to capture the holistic and permutation-invariant nature of mixture perception. This synergistic mechanism not only significantly enhances predictive performance but also provides a physically interpretable geometric framework for the structured organization of the odor space. Although this study represents a breakthrough in static structural representation, future research should further explore the temporal dynamics of olfactory stimuli and the nonlinear perceptual shifts driven by concentration variations.

References

Ann-Sophie Barwich and Elisabeth A Lloyd. More than meets the ai: The possibilities and limits of machine learning in olfaction. *Frontiers in neuroscience*, 16:981294, 2022.

Caroline Bushdid, Marcelo O Magnasco, Leslie B Vosshall, and Andreas Keller. Humans can discriminate more than 1 trillion olfactory stimuli. *Science*, 343(6177):1370–1372, 2014.

Dong Chen, Kaifu Gao, Duc Duy Nguyen, Xin Chen, Yi Jiang, Guo-Wei Wei, and Feng Pan. Algebraic graph-assisted bidirectional transformers for molecular property prediction. *Nature communications*, 12(1):3521, 2021.

Federica Chiera, Giosuè Costa, Stefano Alcaro, and Anna Artese. An overview on olfaction in the biological, analytical, computational, and machine learning fields. *Archiv der Pharmazie*, 358(1): e2400414, 2025.

Seyone Chithrananda, Judith Amores, and Kevin K Yang. Mapping the combinatorial coding between olfactory receptors and perception with deep learning. *bioRxiv*, pp. 2024–09, 2024.

Gülce Nazlı Dikeçligil and Jay A Gottfried. What does the human olfactory system do, and how does it do it? *Annual review of psychology*, 75(1):155–181, 2024.

A. D'Amico, C. Di Natale, R. Paolesse, A. Macagnano, E. Martinelli, G. Pennazza, M. Santonico, M. Bernabei, C. Roscioni, G. Galluccio, R. Bono, E. Finazzi Agrò, and S. Rullo. Olfactory systems for medical applications. *Sensors and Actuators B: Chemical*, 130(1):458–465, 2008. ISSN 0925-4005. doi: https://doi.org/10.1016/j.snb.2007.09.044. Proceedings of the Eleventh International Meeting on Chemical Sensors IMCS-11.

Fragrance Flavor. Food, and cosmetics ingredients information. *The Good Scents Company*, 2018.

Zhihao Hao, Haisheng Li, Jianhua Guo, and Yong Xu. Advances in artificial intelligence for olfaction and gustation: A comprehensive review. *Artificial Intelligence Review*, 58(10):306, 2025.

Lei Huang, Tingyang Xu, Yang Yu, Peilin Zhao, Xingjian Chen, Jing Han, Zhi Xie, Hailong Li, Wenge Zhong, Ka-Chun Wong, et al. A dual diffusion model enables 3d molecule generation and lead optimization based on target pockets. *Nature Communications*, 15(1):2657, 2024.

Tim Hulsen, Saumya S Jamuar, Alan R Moody, Jason H Karnes, Orsolya Varga, Stine Hedensted, Roberto Spreafico, David A Hafler, and Eoin F McKinney. From big data to precision medicine. *Frontiers in medicine*, 6:34, 2019.

Rui Jiao, Jiaqi Han, Wenbing Huang, Yu Rong, and Yang Liu. Energy-motivated equivariant pretraining for 3d molecular graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pp. 8096–8104, 2023.

Andreas Keller, Richard C Gerkin, Yuanfang Guan, Amit Dhurandhar, Gabor Turu, Bence Szalai, Joel D Mainland, Yusuke Ihara, Chung Wen Yu, Russ Wolfinger, et al. Predicting human olfactory perception from chemical features of odor molecules. *Science*, 355(6327):820–826, 2017.

Brian K Lee, Emily J Mayhew, Benjamin Sanchez-Lengeling, Jennifer N Wei, Wesley W Qian, Kelsie A Little, Matthew Andres, Britney B Nguyen, Theresa Moloy, Jacob Yasonik, et al. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006, 2023.

Shengchao Liu, Weitao Du, Zhi-Ming Ma, Hongyu Guo, and Jian Tang. A group symmetric stochastic differential equation model for molecule multi-modal pretraining. In *International Conference on Machine Learning*, pp. 21497–21526. PMLR, 2023.

Chengqiang Lu, Qi Liu, Chao Wang, Zhenya Huang, Peize Lin, and Lixin He. Molecular property prediction: A multilevel quantum interactions modeling perspective. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pp. 1052–1060, 2019.

Johannes Polster and Peter Schieberle. Structure–odor correlations in homologous series of mercaptoalkanols. *Journal of Agricultural and Food Chemistry*, 65(21):4329–4340, 2017.

Aharon Ravia, Kobi Snitz, Danielle Honigstein, Maya Finkel, Rotem Zirler, Ofer Perl, Lavi Secundo, Christophe Laudamiel, David Harel, and Noam Sobel. A measure of smell enables the creation of olfactory metamers. *Nature*, 588(7836):118–123, 2020.

Vivek Sagar, Laura K Shanahan, Christina M Zelano, Jay A Gottfried, and Thorsten Kahnt. High-precision mapping reveals the structure of odor coding in the human brain. *Nature neuroscience*, 26(9):1595–1602, 2023.

Pinaki Saha, Mrityunjay Sharma, Sarabeshwar Balaji, Aryan Amit Barsainyan, Ritesh Kumar, Volker Steuber, and Michael Schmuker. Dense sense: a novel approach utilizing electron density augmented machine learning paradigm to understand the complex odour landscape. *Digital Discovery*, 4(11):3339–3350, 2025. ISSN 2635-098X.

Chao Shang, Qinqing Liu, Ko-Shin Chen, Jiangwen Sun, Jin Lu, Jinfeng Yi, and Jinbo Bi. Edge attention-based multi-relational graph convolutional networks. *arXiv preprint arXiv:1802.04944*, 2, 2018.

Anju Sharma, Rajnish Kumar, Shabnam Ranjta, and Pritish Kumar Varadwaj. Smiles to smell: decoding the structure–odor relationship of chemical compounds using the deep neural network approach. *Journal of Chemical Information and Modeling*, 61(2):676–688, 2021.

Sergey Shuvaev, Khue Tran, Khristina Samoilova, Cyrille Mascart, and Alexei Koulakov. Deepnose: An equivariant convolutional neural network predictive of human olfactory percepts. In *2024 58th Asilomar Conference on Signals, Systems, and Computers*, pp. 536–543. IEEE, 2024.

Kobi Snitz, Adi Yablonka, Tali Weiss, Idan Frumin, Rehan M Khan, and Noam Sobel. Predicting odor perceptual similarity from odor structure. *PLoS computational biology*, 9(9):e1003184, 2013.

Hannes Stärk, Dominique Beaini, Gabriele Corso, Prudencio Tossou, Christian Dallago, Stephan Günnemann, and Pietro Liò. 3d infomax improves gnns for molecular property prediction. In *International Conference on Machine Learning*, pp. 20479–20502. PMLR, 2022.

Farzaneh Taleb, Miguel Vasco, Antonio Ribeiro, Mårten Björkman, and Danica Kragic. Can transformers smell like humans? *Advances in Neural Information Processing Systems*, 37:72032–72060, 2024.

Gary Tom, Cher Tian Ser, Ella M Rajaonson, Stanley Lo, Hyun Suk Park, Brian K Lee, and Benjamin Sanchez-Lengeling. Does this smell the same? learning representations of olfactory mixtures using inductive biases. *Machine Learning: Science and Technology*, 6(3):035063, 2025.

Casey Trimmer, Andreas Keller, Nicolle R Murphy, Lindsey L Snyder, Jason R Willer, Maira H Nagai, Nicholas Katsanis, Leslie B Vosshall, Hiroaki Matsunami, and Joel D Mainland. Genetic variation across the human olfactory receptor repertoire alters odor perception. *Proceedings of the National Academy of Sciences*, 116(19):9475–9480, 2019.

E Vigneau, P Courcoux, Ronan Symoneaux, L Guérin, and Angélique Villière. Random forests: A machine learning methodology to highlight the volatile organic compounds involved in olfactory perception. *Food Quality and Preference*, 68:135–145, 2018.

Yuyang Wang, Jianren Wang, Zhonglin Cao, and Amir Barati Farimani. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence*, 4(3):279–287, 2022.

Mengji Zhang, Yusuke Hiki, Akira Funahashi, and Tetsuya J. Kobayashi. A deep position-encoding model for predicting olfactory perception from molecular structures and electrostatics. *npj Systems Biology and Applications*, 10(1):76, 2024a. doi: 10.1038/s41540-024-00401-0.

Mengji Zhang, Yusuke Hiki, Akira Funahashi, and Tetsuya J Kobayashi. A deep position-encoding model for predicting olfactory perception from molecular structures and electrostatics. *npj Systems Biology and Applications*, 10(1):76, 2024b.

Odin Zhang, Jintu Zhang, Jieyu Jin, Xujun Zhang, RenLing Hu, Chao Shen, Hanqun Cao, Hongyan Du, Yu Kang, Yafeng Deng, et al. Resgen is a pocket-aware 3d molecular generation model based on parallel multiscale modelling. *Nature Machine Intelligence*, 5(9):1020–1030, 2023.

Jinhua Zhu, Yingce Xia, Lijun Wu, Shufang Xie, Tao Qin, Wengang Zhou, Houqiang Li, and Tie-Yan Liu. Unified 2d and 3d pre-training of molecular representations. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*, pp. 2626–2636, 2022.

A Appendix

A.1 Reproduction of the Olfactory White Phenomenon

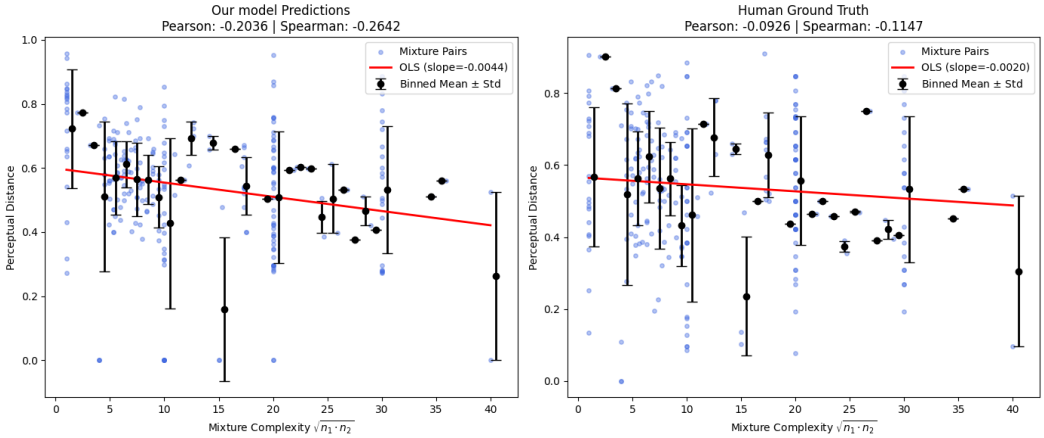


Figure 3: Olfactory White Phenomenon.

The olfactory white phenomenon refers to the perceptual convergence of sufficiently complex odor mixtures. As the number of diverse molecular components in an intensity-balanced mixture increases, the resulting perception progressively loses its distinctive characteristics and converges toward a common, neutral odor, regardless of the specific molecular identities. Analogous to white noise in audition or white light in vision, this phenomenon indicates that the human olfactory system perceives highly complex scent mixtures holistically rather than additively. To examine whether our proposed model captures this macroscopic property of human olfaction, we conduct a reproduction experiment on the mixture datasets.

A representative result evaluated on the held-out test set is shown on the left of Figure 3, with the corresponding human ground truth on the right. Each point denotes a mixture pair, with the horizontal coordinate $\sqrt{n_1 \cdot n_2}$ representing the geometric mean of their component counts (mixture complexity) and the vertical coordinate representing the perceptual distance. The black error bars are obtained by binning pairs into integer-width intervals of the geometric mean, where the centers and bar lengths indicate the within-bin mean and standard deviation, respectively. The red line is an ordinary least squares (OLS) regression fitted to all the scatter points. The statistically significant negative slope confirms that the predicted perceptual distance decreases as mixture complexity increases, a precise signature of *olfactory white*. Notably, our proposed model exhibits an even stronger negative correlation than the ground-truth human judgments do (Pearson: -0.2036 vs. -0.0926; Spearman: -0.2642 vs. -0.1147). This suggests that the model captures the underlying perceptual convergence more consistently than crowdsourced human ratings, which are naturally subject to individual evaluation noise. In conclusion, our model establishes a perceptive structure that is deeply aligned with the macrolevel behavior of the human olfactory system.