NOSE: Neural Olfactory-Semantic Embedding with Tri-Modal Orthogonal Contrastive Learning

Yanyi Su[1,2]*, **Hongshuai Wang**[2], **Zhifeng Gao**[2]†, **Jun Cheng**[1,3,4]†,
[1]State Key Laboratory of Physical Chemistry of Solid Surface,
College of Chemistry and Chemical Engineering, Xiamen University, Xiamen, China
[2]DP Technology,
[3]Laboratory of AI for Electrochemistry (AI4EC),
Tan Kah Kee Innovation Laboratory (IKKEM), Xiamen, China
[4]Institute of Artificial Intelligence, Xiamen University, Xiamen, China

Correspondence: gaozf@dp.tech, chengjun@xmu.edu.cn

Abstract

Olfaction lies at the intersection of chemical structure, neural encoding, and linguistic perception, yet existing representation methods fail to fully capture this pathway. Current approaches typically model only isolated segments of the olfactory pathway, overlooking the complete chain from molecule to receptors to linguistic descriptions. Such fragmentation yields learned embeddings that lack both biological grounding and semantic interpretability. We propose NOSE (Neural Olfactory-Semantic Embedding), a representation learning framework that aligns three modalities along the olfactory pathway: molecular structure, receptor sequence, and natural language description. Rather than simply fusing these signals, we decouple their contributions via orthogonal constraints, preserving the unique encoded information of each modality. To address the sparsity of olfactory language, we introduce a weak positive sample strategy to calibrate semantic similarity, preventing erroneous repulsion of similar odors in the feature space. Extensive experiments demonstrate that NOSE achieves state-of-the-art (SOTA) performance and excellent zero-shot generalization, confirming the strong alignment between its representation space and human olfactory intuition. Code and data are available at `https://github.com/Xianyusyy/NOSE`

1 Introduction

Among the major senses, olfaction is arguably the most challenging to digitize. Vision has pixels and hearing has spectra. These physical quantities maintain stable mappings to perception. But olfaction is different: the same molecule may activate different combinations of receptors, and human descriptions of odors are highly subjective. Olfaction initiates with molecular-receptor binding, propagates through neural signal transduction, and culminates in perceptual formation within the brain (Buck and Axel, 1991; Su et al., 2009; Sobel et al., 1998; Lapid et al., 2011).

To provide context for a broader audience, we briefly introduce the key concepts. In molecular informatics, SMILES (Simplified Molecular-Input Line-Entry System) (Weininger, 1988) is a standard notation that encodes molecular graphs as ASCII strings; for example, `CCO` represents ethanol and `c1ccccc1` represents benzene. Olfactory receptors (ORs) are G-protein-coupled receptor proteins located in the nasal epithelium. The human genome encodes approximately 400 functional ORs (Malnic et al., 1999; Buck, 2004), each responding to specific molecular features, with sequences typically spanning ∼310 amino acids. When an odorant molecule binds to a receptor, it triggers a neural signal cascade that ultimately produces a conscious percept. The central computational challenge in this domain is therefore to predict, given a molecular structure, the perceptual attributes that humans would report, ranging from basic properties such as detection threshold, intensity, and pleasantness, to high-level semantic descriptors such as "floral" or "sweet." Representative inputs for each modality are illustrated in Table 1.

However, existing methods typically model only fragments of this pathway (Jiang et al., 2025; Lee et al., 2023; Chithrananda et al., 2024; Gupta et al.,

*Work done during internship at DP Technology.
†Corresponding author.

Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 19615–19647
July 2-7, 2026

Modality	Input Example
Molecule	`CC1CCCCCCCCCCCCC(=O)C1` (Muscone)
Receptor	`MRENNQSSTLEFILLGVTG...` (OR5A2, 312 amino acids)
Description	"musk; powdery; sweet; floral"

Table 1: Representative input examples for the three pre-training modalities. Molecule and receptor inputs correspond to Muscone and its cognate receptor OR5A2, respectively.

2021). They focus either solely on molecular structure or learn only molecule-description/receptor correspondences, but have never captured the complete chain from molecule to receptor to semantics within a unified framework. A more fundamental issue lies in the task formulation: mainstream methods treat odor prediction as a classification problem, predicting whether a given molecule belongs to "floral" or "fruity". This discretization leads to two consequences. First, it destroys the continuity of odor space. "minty" and "cooling" are highly correlated in human perception, but under the classification framework they are merely two independent labels, leaving the model unable to learn such associations. Similarly, sequence-similar receptors often mediate related odor responses, yet this continuity is also ignored. Second, classification objectives erode the molecular representation itself. When models are forced to fit classification boundaries for odor labels, they often discard information that is useless for classification but crucial for molecular structure. This causes models to perform reasonably on known odor categories but fail to generalize to novel molecules or descriptions.

These observations motivate us to rethink how olfactory representations should be constructed. We propose NOSE (Neural Olfactory-Semantic Embedding), a tri-modal learning framework designed with molecules as the central hub. The core insight stems from a pragmatic observation: although "molecule-receptor-odor description" triplets are extremely scarce, "molecule-receptor" and "molecule-description" bimodal data can be obtained separately. Molecules are the sole intersection of both, serving as a hub to bridge receptor information and odor semantic information into a unified representation space. A question naturally arises: if receptor features and odor semantic features are simultaneously injected into molecular representations, will the three modalities interfere with and overwrite each other? Our solution is orthogonal injection, forcing the two types of features to occupy orthogonal subspaces in the representation space. Receptor information and odor semantic information are superimposed onto molecular representations as mutually independent increments, preserving the integrity of molecular structure while achieving implicit tri-modal alignment. Another practical challenge is the sparsity of odor description data: each molecule is typically annotated with only a few descriptors, while semantically similar words are often treated as irrelevant labels. To address this, we leverage large language models to mine odor semantic proximity relationships among descriptors, expanding isolated labels into continuous odor semantic neighborhoods and mitigating the false negative problem in contrastive learning.

Our contributions span three aspects:

(1) Data Infrastructure: We integrate and curate multi-source odor description and receptor data, constructing the first large-scale pre-training dataset supporting tri-modal learning along with accompanying evaluation benchmarks.

(2) Representation Learning Framework: We propose an orthogonal feature injection mechanism that achieves global alignment and decoupled representations of molecules, receptors, and semantics without relying on triplet annotations.

(3) Semantic Topology Preservation: We design a LLM-enhanced contrastive learning strategy that transforms the discrete odor label space into a continuous odor semantic manifold, mitigating false negative issues caused by label sparsity.

2 Related Work

Predicting odor perception is extremely challenging because it relies on multiple interacting information sources. Early work focused on quantitative structure-activity relationships (QSAR), attempting to predict odor perception solely from molecular structure (Jiang et al., 2025; Sharma et al., 2025; Taleb et al., 2024; Shin et al., 2023; Ravia et al., 2020; Tran et al., 2019; Keller et al., 2017). However, structure-odor relationships are inherently nonlinear (Sell, 2006): minor structural changes can cause dramatic perceptual shifts, while structurally dissimilar molecules may smell alike.

To address this, recent research has begun incorporating auxiliary modalities (odor semantics (Tom et al., 2025; Lee et al., 2023), receptor information (McConachie et al., 2025; Wakutsu and

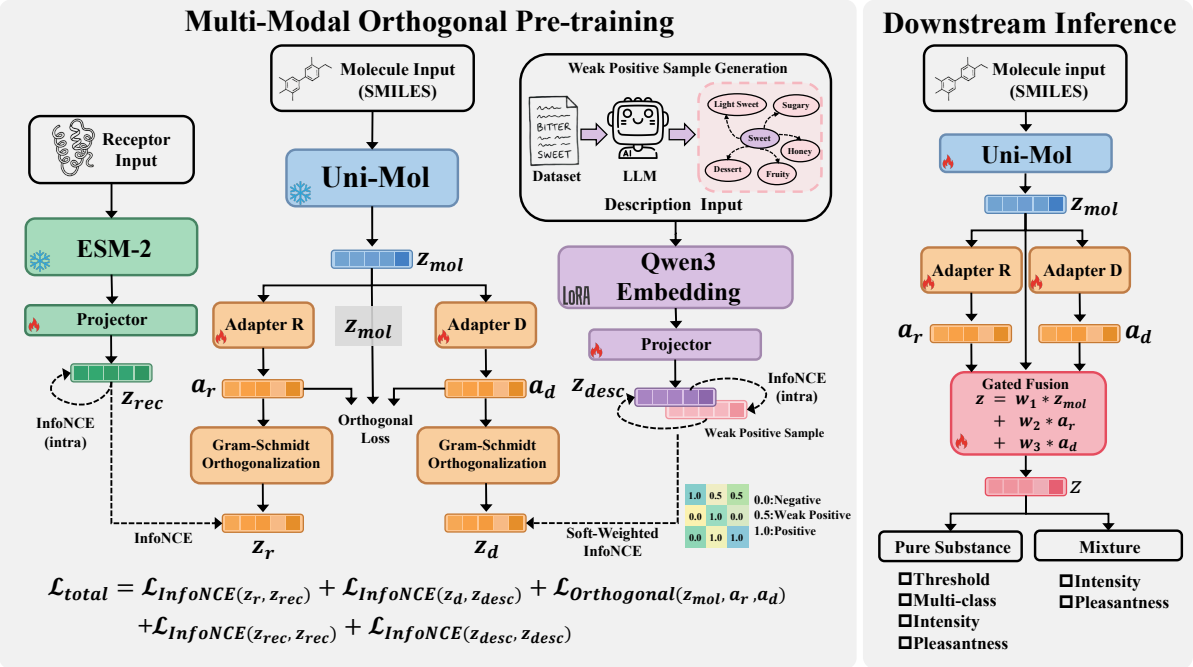

Figure 1: Subscripts r and d denote receptor and description modalities, respectively. **(Left)** Multimodal Orthogonal Pre-training: Molecular representations z_{mol} are extracted by a frozen Uni-Mol encoder, receptor embeddings z_{rec} are obtained from ESM-2 with a trainable projection layer, and odor semantic descriptions z_{desc} are extracted by LoRA-finetuned Qwen3 Embedding after LLM-based weak positive augmentation. The molecular embedding is decomposed via dual adapters into a receptor-aligned component (a_r) and a description-aligned component (a_d), which are then orthogonalized through Gram-Schmidt to yield z_r and z_d. Training objectives include: receptor-molecule InfoNCE loss, description-molecule soft-weighted InfoNCE loss (with positive/weak-positive/negative sample weights of 1.0/0.5/0.0), intra-modal InfoNCE loss, and orthogonality constraints among z_{mol}, a_r, and a_d. **(Right)** Downstream Inference: Inference requires only the molecular encoder and adapters, with the final representation $Z = w_1 \cdot z_{mol} + w_2 \cdot a_r + w_3 \cdot a_d$, supporting both pure substance and mixture perception tasks.

Kaneko, 2025; Chithrananda et al., 2024; Gupta et al., 2021)). However, these methods rely on strongly supervised paradigms, reducing representation learning to multi-label classification or binary receptor matching. This forces models to learn classification boundaries rather than continuous representations, causing molecular, receptor, and semantic features to become entangled in the latent space. Beyond these limitations, even with auxiliary modalities, these methods do not cover the complete olfactory perception pathway from molecules through receptors to semantic descriptions.

Another limitation of existing methods is the neglect of intra-modal topological structure. On the semantic side, existing methods treat odor descriptors as discrete labels, ignoring the continuous semantic relationships among descriptors. For instance, "lemon" and "sour" should be adjacent in odor space. On the receptor side, binary binding prediction ignores the evolutionary homology and family hierarchy among receptor sequences. Such discretization discards intra-modal relationships as core priors, leaving models to learn only isolated mapping rules and unable to construct continuous olfactory spaces consistent with semantic intuition and biological mechanisms.

3 Methodology

3.1 Dataset

3.1.1 Downstream task dataset

Olfactory perception is a multi-level cognitive process. To comprehensively evaluate the quality of olfactory representations, we design an evaluation benchmark spanning three levels: **(1) Basic Perceptual Attribute Prediction**: The most fundamental human perception of odors include "can it be smelled" (threshold), "how strong is it" (intensity), and "is it pleasant" (pleasantness). **(2) Semantic Description Prediction**: This level ex-

Category	Task	Data Format	Source (Size)
Pretraining Corpora	Contrastive	SMILES–Receptor Sequence pairs	Lalis et al., etc. (3,877)
	Contrastive	SMILES–Odor Description pairs	Arctander, etc. (88,512)
		SMILES–Odor Description pairs	LLM-augmented* (2,567,558)
Basic Perception	Threshold	SMILES $\rightarrow [-2.78, 6.11]$	Abraham et al. (268)
	Pleasantness	SMILES $\rightarrow [0, 100]$	Keller and Vosshall (474)
		SMILES $\rightarrow [-1, 1]$	Sagar et al. (160)
	Intensity	SMILES $\rightarrow [0, 100]$	Ravia et al. (248)
		SMILES $\rightarrow [0, 100]$	Keller and Vosshall (474)
		SMILES $\rightarrow [-1, 1]$	Sagar et al. (160)
Semantic Description	Descriptor Classification	SMILES $\rightarrow$ 138 classes	Tom et al. (4,814)
	Descriptor Strength	SMILES $\rightarrow [0, 100]$ (21-dim)	Keller and Vosshall (474)
		SMILES $\rightarrow [-1, 1]$ (15-dim)	Sagar et al. (160)
Mixture Perception	Mixture Intensity	Mixture $\rightarrow [0, 10]$	Ma et al. (6,660)
	Mixture Pleasantness	Mixture $\rightarrow [0, 10]$	Ma et al. (6,660)

Table 2: Overview of pretraining corpora and downstream evaluation tasks. Descriptor Classification is a multi-label classification task, while Descriptor Strength involves multi-dimensional regression. *We leverage an LLM to scale up the pretraining corpora by orders of magnitude.

amines whether the model can capture the mapping between molecular structure and high-level semantic concepts such as "creamy" and "grassy." **(3) Mixture Perception Prediction**: Targeting the ubiquitous mixed odors in real-world scenarios, this level evaluates the model's capability to capture complex nonlinear intermolecular interactions such as masking and synergy. Based on this framework, we integrate six public datasets: Abraham et al. (2012), GS-LF (Lee et al., 2023; Sanchez-Lengeling et al., 2019; Tom et al., 2025), Ravia et al. (2020), Keller and Vosshall (2016), Sagar et al. (2023), and Ma et al. (2021), constructing 11 downstream tasks (see Table 2 and Section A.1 for details).

3.1.2 Pre-training dataset

Receptor Data: We integrate multi-source olfactory receptor data including Pred-O3 (Ollitrault et al., 2024), OlfactionBase (Sharma et al., 2022), M2OR (Lalis et al., 2024), as well as literature-derived data from Mainland et al. (2015) and Ahmed et al. (2018). We construct an olfactory receptor-ligand interaction dataset encompassing extensive biological specificity. Data cleaning and standardization procedures are conducted. Non-human source data and entries that cannot be matched to standard sequences are rigorously excluded. All receptor IDs are converted to protein sequences. For entries with failed queries, manual tracing and sequence completion are performed. After removing duplicate entries across datasets, we construct a SMILES-receptor dataset containing 3,877 activation pairs.

Description Data: We integrate 11 mainstream data sources spanning public databases, academic literature, and industrial catalogs to construct a large-scale comprehensive semantic dataset. Specifically, this encompasses public resources including Arctander (2017), The Good Scents Company (2025), OlfactionBase (Sharma et al., 2022), Sanchez-Lengeling et al. (2019), Flavornet (Acree, 2004), FlavorDB (Garg et al., 2018), AromaDb (Kumar et al., 2018), deep learning benchmark datasets (Sharma et al., 2021), as well as industrial catalogs such as International Fragrance Association (IFRA) (2025) and Sigma-Aldrich (2025). Additionally, we introduce odorless molecules from Mayhew et al. (2022) as negative sample supplements.

The total volume of raw data amounts to approximately 140,000 entries. We implement a rigorous cleaning pipeline, including removal of invalid molecular structures, unification of ambiguous semantic labels, and correction of distributional biases in specific categories. This process yields 88,512 valid data pairs. We retain differentiated descriptions from multiple sources for the same molecule to reflect the inherent subjective diversity of odor perception. To address the challenge

of descriptor sparsity, we leverage the DeepSeek (Guo et al., 2025) model to construct an odor semantic space and mine weak positive samples for all the 1,086 odor descriptors. The final dataset encompasses 9,513 unique molecules, and after semantic augmentation, the total scale is significantly expanded to 2,567,558 SMILES-olfactory descriptor pairs. Details of data cleaning, SMILES standardization, and molecular overlap analysis are provided in Appendix A.2.

3.2 Encoder Selection

Molecular Encoder: To capture the spatial conformations critical for receptor binding, we employ Uni-Mol (Zhou et al., 2023) as the molecular encoder. Pre-trained on 209 million molecular conformations in 3D, this model is capable of modeling stereochemical geometric information underlying molecule-receptor interactions. **Receptor Encoder**: We utilize the ESM-2(650M) (Lin et al., 2023) to extract receptor sequence features. Pre-trained on large-scale protein databases, this model can accurately capture the implicit structural patterns from one-dimensional amino acid sequences. **Odor Descriptor Encoder**: To capture the rich semantics of odor descriptors, we adopt the Qwen3 Embedding (Zhang et al., 2025) model (8B). This model possesses strong generalization capabilities for general text and provides robust representations for subsequent injection of olfactory domain knowledge.

3.3 The NOSE Framework

Given that Uni-Mol and ESM-2 possess native representation capabilities for molecular and receptor features, we freeze their parameters to avoid disrupting their high-quality domain distributions. To address the misalignment between generic semantics and olfactory space distribution in Qwen3 Embedding (Kurfalı et al., 2025; Zhong et al., 2024) (see section A.12 for details), we employ LoRA (Hu et al., 2022) for adaptive fine-tuning. This guides the model's representation manifold to migrate from generic contexts to the olfactory perception domain, thereby achieving precise cross-modal semantic alignment.

3.3.1 Projection Heads and Deep Adapters

For ESM-2 and Qwen3 Embedding, we employ standard nonlinear projection heads. This module follows the design of SimCLR (Chen et al., 2020). For Uni-Mol, we design a deep projection adapter based on ResMLP (Touvron et al., 2022), adopting the Pre-LN (Xiong et al., 2020) structure. Furthermore, considering the significant disparity in data scale between protein sequences and odor descriptions, we adopt differentiated configurations for adapters of these two modalities. The description adapter employs a high-capacity 12-layer inverted bottleneck structure to fit the rich textual data, whereas the receptor adapter introduces a bottleneck structure with high dropout rates to prevent overfitting on sparse receptor data.

3.3.2 Orthogonal Mechanism

To prevent feature redundancy during multi-modal contrastive learning, we employ a combined strategy of geometric decoupling and optimization regularization, inspired by recent advances in multi-modal disentanglement (Hazarika et al., 2020; Liu et al., 2023; Liang et al., 2023).

Geometric Decoupling (Hard Orthogonalization) We leverage Gram-Schmidt orthogonalization to project the adapter's raw output a_{adapter} (e.g., a_{r} or a_{d}) onto the orthogonal complement space of z_{mol}. This geometric operation guarantees that the injected features are linearly independent of molecular structure, forcing the adapter to capture modality-specific increments rather than redundant copies of molecular information. Formally,

$$z_{\text{adapter}} = a_{\text{adapter}} - \frac{a_{\text{adapter}} \cdot z_{\text{mol}}}{\|z_{\text{mol}}\|^2 + \epsilon} z_{\text{mol}} \quad (1)$$

Optimization Regularization (Soft Orthogonalization) While the hard projection above provides a geometric guarantee per sample, it is a unidirectional operation that does not address the interdependency between the receptor branch a_{r} and the description branch a_{d}. We therefore introduce a soft orthogonality loss as a gradient-level regularizer, encouraging all three subspaces to remain mutually decorrelated throughout training. By minimizing the pairwise cosine similarity across subspaces, this loss drives the adapters to learn intrinsically differentiated representations rather than relying solely on the explicit projection.

$$\mathcal{L}_{orth} = \sum_{(i,j)\in\mathcal{S}, i\neq j} \left\| \frac{z_i}{\|z_i\|} \cdot \frac{z_j}{\|z_j\|} \right\|^2 \quad (2)$$

where the feature set $\mathcal{S} = \{z_{mol}, a_r, a_d\}$.

3.3.3 Multi-modal Contrastive Learning

When aligning molecular structures with odor descriptions, since the data originates from multiple

	Thresholds	Pleasantness		Intensity		
Dataset	Abraham	Keller	Sagar	Keller	Sagar	Ravia
GIN	0.72(0.02)	0.30(0.04)	0.08(0.06)	0.06(0.14)	-0.06(0.27)	0.19(0.11)
GCN	0.22(0.09)	0.43(0.01)	0.12(0.27)	0.11(0.05)	0.30(0.04)	0.19(0.01)
Morgan	0.60(0.18)	0.51(0.03)	0.21(0.06)	0.05(0.02)	0.23(0.10)	0.10(0.19)
AttentiveFP	0.59(0.35)	0.52(0.05)	0.23(0.05)	0.33(0.01)	0.23(0.05)	0.47(0.00)
Uni-Mol	0.78(0.05)	0.68(0.02)	0.14(0.34)	0.27(0.03)	0.37(0.05)	0.31(0.02)
POM	0.79(0.02)	0.68(0.03)	0.26(0.02)	0.32(0.07)	0.33(0.06)	0.32(0.06)
ChemBERTa	0.81(0.02)	0.65(0.04)	0.15(0.12)	0.39(0.07)	0.45(0.08)	0.47(0.06)
NOSE	**0.84(0.03)**	**0.71(0.05)**	**0.40(0.02)**	**0.42(0.01)**	**0.47(0.07)**	**0.49(0.03)**

Table 3: Performance comparison on basic perceptual attribute prediction tasks (Pearson ↑).

	Descriptor Classification			Descriptor Strength			
Dataset	GS-LF			Keller		Sagar	
Metric	AUC ↑	AUPRC ↑	MCC ↑	MAE ↓	Pearson ↑	MAE ↓	Pearson ↑
GIN	0.856(0.001)	0.270(0.008)	0.064(0.007)	6.528(0.139)	0.128(0.007)	0.358(0.005)	-0.034(0.016)
GCN	0.858(0.001)	0.279(0.001)	0.084(0.003)	6.247(0.031)	0.171(0.006)	0.407(0.098)	0.108(0.085)
Morgan	0.850(0.002)	0.320(0.002)	0.222(0.023)	6.483(0.015)	0.244(0.023)	0.395(0.018)	0.008(0.076)
POM	0.868(0.002)	0.336(0.005)	0.233(0.023)	6.304(0.258)	0.314(0.039)	0.405(0.113)	0.065(0.062)
AttentiveFP	0.867(0.004)	0.328(0.012)	0.203(0.026)	6.232(0.647)	0.327(0.065)	0.359(0.012)	0.011(0.013)
Uni-Mol	0.873(0.001)	0.347(0.005)	0.262(0.020)	6.741(0.209)	0.330(0.050)	0.355(0.009)	0.116(0.042)
ChemBERTa	0.875(0.001)	0.342(0.005)	0.240(0.016)	6.110(0.255)	0.330(0.089)	0.376(0.018)	0.105(0.051)
NOSE	**0.876(0.001)**	**0.351(0.002)**	**0.268(0.010)**	**5.862(0.225)**	**0.348(0.060)**	**0.343(0.017)**	**0.123(0.068)**

Table 4: Performance comparison on semantic description prediction tasks.

datasets, different descriptors often exhibit similar semantics (e.g., "sweet" and "honey"). If the standard contrastive learning paradigm is adopted, these semantically similar descriptors would be incorrectly treated as negative samples, causing the model to erroneously push apart samples with similar odors. We employ DeepSeek to match each descriptor with semantically similar olfactory terms from the dataset as weak positive samples (generation details in Appendix A.2.4). These weak positive samples are not only utilized for cross-modal alignment between molecular structures and odor descriptions, but also incorporated into the intra-modal contrastive learning within the odor description modality itself.

The contrastive learning component is based on the CLIP (Radford et al., 2021) paradigm and employs a symmetric bidirectional InfoNCE loss. To address the inherent many-to-many mapping relationships among molecules, receptors, and odor descriptions, we reformulate the definitions of positive and negative samples. In cross-modal alignment, besides the sample itself, all entities sharing the same molecular structures with the anchor are also treated as positive samples. To exploit the feature potential within individual modalities, we introduce intra-modal contrastive learning (Yuan et al., 2021). In the intra-modal space, positive samples are similarly defined as the sample itself and the set of samples belonging to the same molecular structures. Notably, for the odor description modality, weak positive samples generated by LLMs are also incorporated into the positive sample set. Considering that weak positive samples are semantically similar to but not fully equivalent to the anchor, we draw upon the soft contrastive learning (Zhou et al., 2021) framework and define a weight function w_{ij} to achieve smooth supervision:

$$w_{ij} = \begin{cases} 1.0 & \text{Positive Sample} \\ 0.5 & \text{Weak Positive Sample} \\ 0 & \text{Negative Sample} \end{cases} \tag{3}$$

The contrastive learning loss is formulated as:

$$\mathcal{L}^{a\to b} = -\frac{1}{|\mathcal{V}|}\sum_{i\in\mathcal{V}} \frac{\sum_{j=1}^{N} w_{ij}\cdot \log \frac{\exp(s_{ij}/\tau)}{\sum_{k=1}^{N}\exp(s_{ik}/\tau)}}{\sum_{j=1}^{N} w_{ij}} \tag{4}$$

where s_{ij} denotes the cosine similarity between samples i and j, τ is the temperature parameter, and $\mathcal{V}$ represents the set of anchors with at least one positive sample. The total contrastive learning

Metric	Mixture Pleasantness				Mixture Intensity			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
GCN	-14.30(9.52)	3.69(1.72)	0.33(0.42)	17.48(10.88)	-0.46(0.50)	0.52(0.09)	0.49(0.07)	0.40(0.14)
GIN	-0.10(0.12)	0.88(0.06)	0.47(0.03)	1.25(0.14)	-2.82(0.30)	0.86(0.02)	0.01(0.14)	1.06(0.08)
Morgan	0.50(0.19)	0.60(0.11)	0.81(0.01)	0.57(0.21)	0.14(0.31)	0.40(0.08)	0.58(0.06)	0.24(0.08)
AttentiveFP	0.49(0.07)	0.62(0.05)	0.74(0.02)	0.58(0.08)	0.26(0.06)	0.36(0.01)	0.60(0.03)	0.20(0.02)
POM	0.56(0.03)	0.56(0.00)	0.77(0.02)	0.51(0.03)	0.35(0.06)	0.36(0.02)	0.60(0.03)	0.18(0.02)
ChemBERTa	0.45(0.07)	0.64(0.06)	0.79(0.01)	0.63(0.08)	0.33(0.02)	0.35(0.00)	0.62(0.00)	0.19(0.01)
Uni-Mol	0.51(0.05)	0.61(0.02)	0.80(0.03)	0.56(0.05)	0.32(0.11)	0.36(0.03)	0.64(0.07)	0.19(0.03)
NOSE	**0.64(0.05)**	**0.53(0.03)**	**0.85(0.01)**	**0.42(0.05)**	**0.39(0.05)**	**0.33(0.02)**	**0.66(0.03)**	**0.17(0.01)**

Table 5: Performance comparison on mixture perception prediction tasks.

loss is:

$$\mathcal{L}_{\text{a-b}} = \mathcal{L}_{\text{inter}}^{\text{a}\to\text{b}} + \mathcal{L}_{\text{inter}}^{\text{b}\to\text{a}} + \mathcal{L}_{\text{intra}}^{\text{a}\to\text{a}} + \mathcal{L}_{\text{intra}}^{\text{b}\to\text{b}} \quad (5)$$

3.3.4 Optimization Objective

The final training objective of the model comprises cross-modal alignment losses and orthogonality constraints, jointly optimized through weighted summation:

$$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}_{\text{mol-desc}} + \lambda_2 \cdot \mathcal{L}_{\text{mol-rec}} + \lambda_3 \cdot \mathcal{L}_{orth} \quad (6)$$

4 Experiments

4.1 Downstream Task

Baseline: Given the absence of widely recognized dedicated benchmarks in olfactory representation learning, we adopt a hierarchical comparison strategy to systematically select baseline models: Morgan Fingerprint (Rogers and Hahn, 2010), GCN (Kipf, 2016), GIN (Xu et al., 2018), AttentiveFP (Xiong et al., 2019), POM (Lee et al., 2023), ChemBERTa (Chithrananda et al., 2020), Uni-Mol (Zhou et al., 2023). The baseline design follows three progressive dimensions: (1) representation scope (local $\rightarrow$ global), (2) learning paradigm (fixed encoding $\rightarrow$ supervised learning $\rightarrow$ pre-training), and (3) domain adaptation (general-purpose $\rightarrow$ olfaction-specific). This enables us to disentangle the contributions of different factors to performance, thereby establishing the performance boundaries of existing general-purpose molecular models on olfactory perception tasks.

Experimental Setup: To ensure reproducibility and robustness, all dataset splits are fixed with random seed (42). Final results are reported as the mean and standard deviation across three independent experiments (seeds: 42-44) on the test set. For the heterogeneous features output by NOSE, we employ a gating mechanism for adaptive fusion. This design introduces only three learnable parameters, ensuring fairness in comparison with baselines. The evaluation benchmark covers three dimensions: basic perception (Table 3), semantic description (Table 4), and mixture prediction (Table 5). Due to space constraints, detailed experimental settings, comparative experiments, ablation studies, and zero-shot generalization results are provided in the appendix (sections A.3, A.9, A.10, A.13 and A.14).

NOSE achieves SOTA on all key metrics, demonstrating the high generalizability of its learned representations. Even on the highly challenging mixture tasks (encoding details in Appendix A.2.3), NOSE maintains excellent performance ($R^2 > 0.6$), indicating that the model effectively captures transferable essential olfactory features. Comparisons with Uni-Mol and ChemBERTa further confirm that pre-trained architectures specifically designed for olfactory perception hold advantages over general-purpose chemical pre-trained models.

4.2 Ablation Experiment

To validate the core components, we conduct ablation experiments using the average rank across all downstream tasks and evaluation metrics (lower is better) as the criterion (Table 6). The results demonstrate that: (1) Tri-modal fusion significantly outperforms uni-modal baselines, confirming that olfactory perception relies on the synergy among molecular structures, receptors, and odor semantics. (2) The combination of soft and hard orthogonality achieves optimal performance, balancing incremental extraction with information preservation to enable deep feature decoupling. (3) The asymmetric adapter design outperforms symmetric configurations, effectively accommodating the disparity in

data scales between the two modalities. (4) Intra-modal contrastive learning is highly coupled with weak positive samples. Their combination ensures both discriminability and continuity, preventing representation space degradation (clustering analysis in Appendix A.8). We further investigate alternative decoupling strategies (dimension splitting vs. orthogonal injection) in Appendix A.6; PCA visualizations confirming the decoupling property are shown in Appendix A.11. Training stability analyses, including cross-seed variance and fusion weight consistency, are provided in Appendix A.7.

Settings	Variant	Avg Rank
Multi-modality	Molecule only	3.3226
	Receptor only	2.8065
	Description only	2.7742
	Tri-modal (ours)	**1.0968**
Orthogonal Module	Hard + Soft λ=0.5	5.1613
	Hard + Soft λ=0.1	5.0645
	No Orthogonal (baseline)	4.5484
	Only Hard	4.1290
	Only Soft λ=2.0	4.0968
	Hard + Soft λ=1.0	3.9355
	Hard + Soft λ=2.0 (ours)	**1.0645**
Adapter Capacity	Desc: 10.00M Rec: 10.00M	2.5161
	Desc: 29.17M Rec: 10.00M	2.4839
	Desc: 76.70M Rec: 4.49M (ours)	**1.0000**
Contrastive Strategy	Inter + Weak	2.9677
	Inter (baseline)	2.8387
	Inter + Intra	2.8065
	Inter + Intra + Weak (ours)	**1.3871**

Table 6: Ablation study results.

4.3 Retrieval Evaluation

To validate the effectiveness of NOSE in unifying molecular structures, receptor sequences, and olfactory semantics. We design compositional retrieval and zero-shot cross-modal retrieval tasks. These two tasks evaluate the feature space from two dimensions: the structural integrity of olfactory semantics and the generalizability of olfactory physiological principles. We adopt MRR and Hits@K as evaluation metrics. Due to space constraints, only the main results are presented here (see sections A.13 and A.14 for details).

4.3.1 Compositional Retrieval

Traditional retrieval tasks only examine "point-to-point" mappings, whereas our proposed compositional retrieval operates within the descriptor space, requiring the model to locate answers based on "Anchor + Operation" queries (e.g., $Lemon - Sour \rightarrow [citrus, orange, ...]$). This aims to evaluate whether contrastive learning successfully decouples odor attributes into operable olfactory semantic units. We constructed 31 manually annotated formula-answer pairs to test this capability. As shown in Table 8, the original Qwen3 Embedding model struggles with these compositions (MRR 0.0102). While LoRA pre-training provides preliminary responsiveness, the model incorporating the contrastive learning head achieves optimal performance (MRR 0.2072, Hits@50 100%). This demonstrates that our method constructs a clear topological structure in the feature space, enabling effective handling of logical attribute queries.

4.3.2 Cross-Modal Zero-Shot Retrieval

To evaluate cross-modal generalization, we conducted zero-shot retrieval using molecular SMILES as queries to retrieve odor descriptors (sourced from PubChem) and receptor sequences (from literature). As shown in Table 7, in the Odor Descriptor Retrieval task, the model maintains effective hit rates even under strict zero-shot settings. This indicates that the model has successfully learned the underlying mapping principles between molecular structures and odor descriptions. Regarding Receptor Sequence Retrieval, the model demonstrates strong discriminative power. For activated receptors, Hits@5 reaches 61.5%, accurately placing correct receptor at the top. Notably, for non-activated receptors, the occurrence rate in the Top-200 is 0%. This confirms that NOSE embedding effectively separates non-interacting pairs, ensuring results conform to authentic biophysical principles. Extended cross-modal retrieval experiments using bridge molecules (Appendix A.4) and direct geometric validation of the continuous perceptual space (Appendix A.5) provide further evidence for the quality of the unified tri-modal embedding.

5 Conclusion

We propose NOSE, the first olfactory representation framework that integrates molecular structures, receptor sequences, and subjective descriptions. We collect large-scale bi-paired data and employ LLM augmentation, constructing a continuous feature manifold through soft-hard orthogonal mechanisms and intra-modal contrastive learning, which effectively decouples and injects receptor and semantic information while preserving molecu-

Task	N	Hits@1	Hits@5	Hits@10	Hits@20	Hits@50
Odor Retrieval (Zero-shot)	27	14.8%	33.3%	51.9%	66.7%	85.2%
Odor Retrieval (Strict Zero-shot)	15	6.7%	6.7%	13.3%	33.3%	53.3%
Receptor Retrieval (Activated)	13	0%	61.5%	69.2%	84.6%	84.6%

Task	N	Hits@100	Hits@200	Hits@300	Hits@400	Hits@500
Receptor Retrieval (Inactivated)	5	0%	0%	20%	60%	100%

Table 7: Zero-shot cross-modal retrieval performance (SMILES → descriptors/receptors). **N**: number of test samples. **Hits@k**: percentage of ground-truth targets ranked within top k. **Zero-shot**: pairs unseen during training, though molecules may appear in other pairs. **Strict Zero-shot**: molecules entirely absent from training. For Activated, higher Hits@k is better (target receptor should rank high). For Inactivated, lower Hits@k is better (non-binding receptor should rank low).

	Qwen3 Embedding		
	Original	**LoRA**	**LoRA+Head**
MRR↑	0.0102	0.1857	**0.2072**
Hits@1 (%)↑	0.0	6.5	**6.5**
Hits@5 (%)↑	0.0	25.8	**32.3**
Hits@10 (%)↑	3.2	45.2	**54.8**
Hits@20 (%)↑	3.2	71.0	**90.3**
Hits@50 (%)↑	6.5	93.5	**100.0**

Table 8: Compositional retrieval performance on 31 manually designed arithmetic queries (e.g., *Lemon* − *Sour* → [citrus, orange, ...]). **Original**: Qwen3 Embedding without fine-tuning; **LoRA**: with LoRA fine-tuning; **LoRA+Head**: LoRA combined with contrastive learning projection head. MRR: Mean Reciprocal Rank ($1/\text{rank}$).

lar structure priors. On our newly proposed benchmark, NOSE achieves state-of-the-art performance across multiple downstream tasks. The model's performance in compositional retrieval and zero-shot cross-modal retrieval confirms its ability to learn a tri-modal space aligned with human intuition from disjoint local paired data.

6 Limitations

Our current framework does not explicitly model concentration effects, despite odor perception being inherently concentration-dependent (e.g., indole shifts from floral to fecal at varying levels). This is primarily due to the lack of concentration annotations in most public datasets. However, such information exists in scattered literature. Collecting and incorporating concentration as a conditional or latent variable presents a promising direction for future work.

On the receptor side, the reliance on contrastive learning alone limits representation quality given the sparse receptor-odor pairings. Integrating molecular docking models could provide complementary structural priors to enhance receptor-level embeddings. For the descriptor modality, the observed label inconsistency stems from heterogeneous data sources with differing annotation standards. Explicitly modeling source provenance as a conditioning factor may help disentangle conflicting descriptions while preserving semantic diversity.

7 Ethical Considerations

Data Usage and Licensing Compliance: We utilize publicly available datasets comprising SMILES, human receptor sequences, and olfactory descriptions from established repositories. As the data is anonymized and non-interactive, IRB approval is exempt. This study strictly adheres to the licensing agreements of all utilized assets. Regarding model implementations, Uni-Mol, ESM-2, POM, and ChemBERTa (built upon DeepChem) are utilized under the MIT License, while Qwen3 Embedding is employed under the Apache 2.0 License. With respect to data and tasks, the majority of odor descriptors and receptor data are obtained through the Pyrfume[1] (Hamel et al., 2024) library, distributed under the MIT License. The M2OR dataset is utilized in accordance with the Apache 2.0 License. For the Pred-O3 dataset, released under the CC BY-NC 4.0 License, this study strictly restricts its usage to academic research purposes without any commercial applications, fully complying with the non-commercial use provisions.

Broader Impact: Our work aims to advance olfactory digitization for applications like flavor design and sensory computing. We acknowledge the theoretical risk of misuse for designing hazardous

[1] https://github.com/pyrfume/pyrfume-data

or offensively targeted compounds. However, we believe the scientific value outweighs these risks. We advocate for responsible deployment and emphasize that AI predictions cannot replace rigorous chemical safety testing.

We used Claude and Gemini for coding assistance and language polishing. All outputs were reviewed and verified by the authors, who bear full responsibility for the content of this work.

Acknowledgments

We are grateful for funding support from the National Natural Science Foundation of China (Grants No. 92470201, 22225302, 92461312, 22541204, 22021001), the Fundamental Research Funds for the Central Universities 20720250005, Laboratory of AI for Electrochemistry (AI4EC), IKKEM (Grant Nos. RD2023100101 and RD2022070501).

References

Michael H Abraham, Ricardo Sánchez-Moreno, J Enrique Cometto-Muñiz, and William S Cain. 2012. An algorithm for 353 odor detection thresholds in humans. *Chemical senses*, 37(3):207–218.

T Acree. 2004. Flavornet and human odor space. *http://www.flavornet.org/flavornet.html*.

Lucky Ahmed, Yuetian Zhang, Eric Block, Michael Buehl, Michael J Corr, Rodrigo A Cormanich, Sivaji Gundala, Hiroaki Matsunami, David O'Hagan, Mehmet Ozbil, and others. 2018. Molecular mechanism of activation of human musk receptors or5an1 and or1a1 by (r)-muscone and diverse other musk-smelling compounds. *Proceedings of the National Academy of Sciences*, 115(17):E3950–E3958.

Steffen Arctander. 2017. *Perfume and flavor materials of natural origin*. Lulu.com.

Christian B Billesbølle, Claire A de March, Wijnand JC van der Velden, Ning Ma, Jeevan Tewari, Claudia Llinas Del Torrent, Linus Li, Bryan Faust, Nagarajan Vaidehi, Hiroaki Matsunami, and others. 2023. Structural basis of odorant recognition by a human odorant receptor. *Nature*, 615(7953):742–749.

Linda Buck and Richard Axel. 1991. A novel multigene family may encode odorant receptors: a molecular basis for odor recognition. *Cell*, 65(1):175–187.

Linda B Buck. 2004. Olfactory receptors and odor coding in mammals. *Nutrition reviews*, 62(suppl_3):S184–S188.

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

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. 2020. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR.

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

Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. 2020. Chemberta: large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885*.

Chulwon Choi, Jungnam Bae, Seonghan Kim, Seho Lee, Hyunook Kang, Jinuk Kim, Injin Bang, Kiheon Kim, Won-Ki Huh, Chaok Seok, and others. 2023. Understanding the molecular mechanisms of odorant binding and activation of the human or52 family. *Nature Communications*, 14(1):8105.

Neelansh Garg, Apuroop Sethupathy, Rudraksh Tuwani, Rakhi Nk, Shubham Dokania, Arvind Iyer, Ayushi Gupta, Shubhra Agrawal, Navjot Singh, Shubham Shukla, and others. 2018. Flavordb: a database of flavor molecules. *Nucleic acids research*, 46(D1):D1210–D1216.

Christiane Geithe, Gaby Andersen, Agne Malki, and Dietmar Krautwurst. 2015. A butter aroma recombinate activates human class-i odorant receptors. *Journal of agricultural and food chemistry*, 63(43):9410–9420.

Rosa S Gisladottir, Erna V Ivarsdottir, Agnar Helgason, Lina Jonsson, Nanna K Hannesdottir, Gudrun Rutsdottir, Gudny A Arnadottir, Astros Skuladottir, Benedikt A Jonsson, Gudmundur L Norddahl, and others. 2020. Sequence variants in taar5 and other loci affect human odor perception and naming. *Current Biology*, 30(23):4643–4653.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Peiyi Wang, Qihao Zhu, Runxin Xu, Ruoyu Zhang, Shirong Ma, Xiao Bi, and others. 2025. Deepseek-r1 incentivizes reasoning in llms through reinforcement learning. *Nature*, 645(8081):633–638.

Ria Gupta, Aayushi Mittal, Vishesh Agrawal, Sushant Gupta, Krishan Gupta, Rishi Raj Jain, Prakriti Garg, Sanjay Kumar Mohanty, Riya Sogani, Harshit Singh Chhabra, and others. 2021. Odorify: a conglomerate of artificial intelligence–driven prediction engines for olfactory decoding. *Journal of Biological Chemistry*, 297(2).

Franziska Haag, Antonella Di Pizio, and Dietmar Krautwurst. 2022. The key food odorant receptive range of broadly tuned receptor or2w1. *Food Chemistry*, 375:131680.

Elizabeth A Hamel, Jason B Castro, Travis J Gould, Robert Pellegrino, Zhiwei Liang, Liyah A Coleman, Famesh Patel, Derek S Wallace, Tanushri Bhatnagar, Joel D Mainland, and others. 2024. Pyrfume: A

window to the world's olfactory data. *Scientific Data*, 11(1):1220.

Devamanyu Hazarika, Roger Zimmermann, and Soujanya Poria. 2020. Misa: Modality-invariant and-specific representations for multimodal sentiment analysis. In *Proceedings of the 28th ACM international conference on multimedia*, pages 1122–1131.

Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, Weizhu Chen, and others. 2022. Lora: Low-rank adaptation of large language models. *ICLR*, 1(2):3.

International Fragrance Association (IFRA). 2025. IFRA Fragrance Ingredient Glossary. https://ifrafragrance.org/publications/guidance-reference-document/ifra-fragrance-ingredient-glossary. Accessed: December 10, 2025.

Yongquan Jiang, Xin Xie, Yan Yang, Yuerui Liu, Kuanping Gong, and Tianrui Li. 2025. Dual-branch graph neural network for predicting molecular odors and discovering the relationship between functional groups and odors. *Journal of Computational Chemistry*, 46(6):e70069.

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

Andreas Keller and Leslie B Vosshall. 2016. Olfactory perception of chemically diverse molecules. *BMC neuroscience*, 17(1):55.

TN Kipf. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Yogesh Kumar, Om Prakash, Himanshu Tripathi, Sudeep Tandon, Madan M Gupta, Laiq-Ur Rahman, Raj K Lal, Manoj Semwal, Mahendra Pandurang Darokar, and Feroz Khan. 2018. Aromadb: a database of medicinal and aromatic plant's aroma molecules with phytochemistry and therapeutic potentials. *Frontiers in plant science*, 9:1081.

Murathan Kurfalı, Pawel Herman, Stephen Pierzchajlo, Jonas Olofsson, and Thomas Hörberg. 2025. Representations of smells: The next frontier for language models? *Cognition*, 264:106243.

Maxence Lalis, Matej Hladiš, Samar Abi Khalil, Loïc Briand, Sébastien Fiorucci, and Jérémie Topin. 2024. M2or: a database of olfactory receptor–odorant pairs for understanding the molecular mechanisms of olfaction. *Nucleic Acids Research*, 52(D1):D1370–D1379.

Hadas Lapid, Sagit Shushan, Anton Plotkin, Hillary Voet, Yehudah Roth, Thomas Hummel, Elad Schneidman, and Noam Sobel. 2011. Neural activity at the human olfactory epithelium reflects olfactory perception. *Nature neuroscience*, 14(11):1455–1461.

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, and others. 2023. A principal odor map unifies diverse tasks in olfactory perception. *Science*, 381(6661):999–1006.

Paul Pu Liang, Zihao Deng, Martin Q Ma, James Y Zou, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2023. Factorized contrastive learning: Going beyond multi-view redundancy. *Advances in Neural Information Processing Systems*, 36:32971–32998.

Zeming Lin, Halil Akin, Roshan Rao, Brian Hie, Zhongkai Zhu, Wenting Lu, Nikita Smetanin, Robert Verkuil, Ori Kabeli, Yaniv Shmueli, and others. 2023. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science*, 379(6637):1123–1130.

Shengzhong Liu, Tomoyoshi Kimura, Dongxin Liu, Ruijie Wang, Jinyang Li, Suhas Diggavi, Mani Srivastava, and Tarek Abdelzaher. 2023. Focal: Contrastive learning for multimodal time-series sensing signals in factorized orthogonal latent space. *Advances in Neural Information Processing Systems*, 36:47309–47338.

Yue Ma, Ke Tang, Yan Xu, and Thierry Thomas-Danguin. 2021. A dataset on odor intensity and odor pleasantness of 222 binary mixtures of 72 key food odorants rated by a sensory panel of 30 trained assessors. *Data in brief*, 36:107143.

Joel D Mainland, Yun R Li, Ting Zhou, Wen Ling L Liu, and Hiroaki Matsunami. 2015. Human olfactory receptor responses to odorants. *Scientific data*, 2(1):1–9.

Bettina Malnic, Junzo Hirono, Takaaki Sato, and Linda B Buck. 1999. Combinatorial receptor codes for odors. *Cell*, 96(5):713–723.

Emily J Mayhew, Charles J Arayata, Richard C Gerkin, Brian K Lee, Jonathan M Magill, Lindsey L Snyder, Kelsie A Little, Chung Wen Yu, and Joel D Mainland. 2022. Transport features predict if a molecule is odorous. *Proceedings of the National Academy of Sciences*, 119(15):e2116576119.

Grant D McConachie, Emily Duniec, Florence Guerina, Meg A Younger, and Brian DePasquale. 2025. Low rank adaptation of chemical foundation models generates effective odorant representations. *bioRxiv*, pages 2025–11.

Guillaume Ollitrault, Rayane Achebouche, Antoine Dreux, Samuel Murail, Karine Audouze, Anne Tromelin, and Olivier Taboureau. 2024. Pred-o3, a web server to predict molecules, olfactory receptors and odor relationships. *Nucleic Acids Research*, 52(W1):W507–W512.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, and others. 2021. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

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

David Rogers and Mathew Hahn. 2010. Extended-connectivity fingerprints. *Journal of chemical information and modeling*, 50(5):742–754.

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

Benjamin Sanchez-Lengeling, Jennifer N Wei, Brian K Lee, Richard C Gerkin, Alán Aspuru-Guzik, and Alexander B Wiltschko. 2019. Machine learning for scent: Learning generalizable perceptual representations of small molecules. *arXiv preprint arXiv:1910.10685*.

Charles S Sell. 2006. On the unpredictability of odor. *Angewandte Chemie International Edition*, 45(38):6254–6261.

Anju Sharma, Rajnish Kumar, Shabnam Ranjta, and Pritish Kumar Varadwaj. 2021. 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.

Anju Sharma, Bishal Kumar Saha, Rajnish Kumar, and Pritish Kumar Varadwaj. 2022. Olfactionbase: a repository to explore odors, odorants, olfactory receptors and odorant–receptor interactions. *Nucleic Acids Research*, 50(D1):D678–D686.

Mrityunjay Sharma, Sarabeshwar Balaji, Pinaki Saha, and Ritesh Kumar. 2025. Navigating the fragrance space using graph generative models and predicting odors. *Journal of Chemical Information and Modeling*, 65(10):4818–4832.

Daniel Shin, Gao Pei, Priyadarshini Kumari, and Tarek R Besold. 2023. Optimizing learning across multimodal transfer features for modeling olfactory perception. In *29th ACM SIGKDD Cof. on Knowledge Discovery and Data Mining*.

Sigma-Aldrich. 2025. Sigma-Aldrich Website. https://www.sigmaaldrich.com/SG/en. Accessed: December 10, 2025.

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

Noam Sobel, Vivek Prabhakaran, John E Desmond, Gary H Glover, RL Goode, Edith V Sullivan, and John DE Gabrieli. 1998. Sniffing and smelling: separate subsystems in the human olfactory cortex. *Nature*, 392(6673):282–286.

Chih-Ying Su, Karen Menuz, and John R Carlson. 2009. Olfactory perception: receptors, cells, and circuits. *Cell*, 139(1):45–59.

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

The Good Scents Company. 2025. The Good Scents Company Information System. https://www.thegoodscentscompany.com/. Accessed: December 10, 2025.

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

Hugo Touvron, Piotr Bojanowski, Mathilde Caron, Matthieu Cord, Alaaeldin El-Nouby, Edouard Grave, Gautier Izacard, Armand Joulin, Gabriel Synnaeve, Jakob Verbeek, and others. 2022. Resmlp: Feedforward networks for image classification with data-efficient training. *IEEE transactions on pattern analysis and machine intelligence*, 45(4):5314–5321.

Ngoc Tran, Daniel Kepple, Sergey Shuvaev, and Alexei Koulakov. 2019. Deepnose: Using artificial neural networks to represent the space of odorants. In *International Conference on Machine Learning*, pages 6305–6314. PMLR.

Yuta Wakutsu and Hiromasa Kaneko. 2025. Molecular odor prediction using olfactory receptor information. *Molecular Informatics*, 44(3):e202400274.

Ivonne Wallrabenstein, Jonas Kuklan, Lea Weber, Sandra Zborala, Markus Werner, Janine Altmüller, Christian Becker, Anna Schmidt, Hanns Hatt, Thomas Hummel, and others. 2013. Human trace amine-associated receptor taar5 can be activated by trimethylamine. *PloS one*, 8(2):e54950.

David Weininger. 1988. Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules. *Journal of chemical information and computer sciences*, 28(1):31–36.

Ruibin Xiong, Yunchang Yang, Di He, Kai Zheng, Shuxin Zheng, Chen Xing, Huishuai Zhang, Yanyan Lan, Liwei Wang, and Tieyan Liu. 2020. On layer

normalization in the transformer architecture. In *International conference on machine learning*, pages 10524–10533. PMLR.

Zhaoping Xiong, Dingyan Wang, Xiaohong Liu, Feisheng Zhong, Xiaozhe Wan, Xutong Li, Zhaojun Li, Xiaomin Luo, Kaixian Chen, Hualiang Jiang, and others. 2019. Pushing the boundaries of molecular representation for drug discovery with the graph attention mechanism. *Journal of medicinal chemistry*, 63(16):8749–8760.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2018. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*.

Keiichi Yoshikawa, Jun Deguchi, Jieying Hu, Hsiu-Yi Lu, and Hiroaki Matsunami. 2022. An odorant receptor that senses four classes of musk compounds. *Current Biology*, 32(23):5172–5179.

Xin Yuan, Zhe Lin, Jason Kuen, Jianming Zhang, Yilin Wang, Michael Maire, Ajinkya Kale, and Baldo Faieta. 2021. Multimodal contrastive training for visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6995–7004.

Shitong Zeng, Lili Zhang, Peng Li, Dandan Pu, Yingjie Fu, Ruiyi Zheng, Hui Xi, Kaina Qiao, Dingzhong Wang, Baoguo Sun, and others. 2023. Molecular mechanisms of caramel-like odorant-olfactory receptor interactions based on a computational chemistry approach. *Food Research International*, 171:113063.

Yanzhao Zhang, Mingxin Li, Dingkun Long, Xin Zhang, Huan Lin, Baosong Yang, Pengjun Xie, An Yang, Dayiheng Liu, Junyang Lin, and others. 2025. Qwen3 embedding: Advancing text embedding and reranking through foundation models. *arXiv preprint arXiv:2506.05176*.

Shu Zhong, Zetao Zhou, Christopher Dawes, Giada Brianz, and Marianna Obrist. 2024. Sniff ai: Is my'spicy'your'spicy'? exploring llm's perceptual alignment with human smell experiences. *arXiv preprint arXiv:2411.06950*.

Gengmo Zhou, Zhifeng Gao, Qiankun Ding, Hang Zheng, Hongteng Xu, Zhewei Wei, Linfeng Zhang, and Guolin Ke. 2023. Uni-mol: A universal 3d molecular representation learning framework. In *The eleventh international conference on learning representations*.

Pan Zhou, Caiming Xiong, Xiaotong Yuan, and Steven Chu Hong Hoi. 2021. A theory-driven self-labeling refinement method for contrastive representation learning. *Advances in Neural Information Processing Systems*, 34:6183–6197.

A Appendix

A.1 Downstream task dataset

The Abraham et al. (2012) dataset compiles odor detection threshold (ODT) data primarily measured by Nagata using the Japanese triangle odor bag method, and unifies multi-source data onto a common scale via an indicator variable algorithm. Raw thresholds (ppm) are transformed into log(1/ODT) form, where higher values indicate greater odor potency.

The GS-LF dataset (Lee et al., 2023; Sanchez-Lengeling et al., 2019) integrates expert olfactory annotations from the GoodScents and Leffingwell databases. After curation by Tom et al. (2025) et al. (standardizing SMILES, removing duplicates, inorganics, salts, and samples with anomalous molecular weights), it contains 4,814 pure substance molecules and 138 semantic descriptors (e.g., "creamy," "grassy").

The Ravia et al. (2020) dataset contains psychophysical intensity ratings for 248 undiluted pure substances. Raw ratings are normalized per subject and then averaged.

The Keller and Vosshall (2016) dataset comprises ratings from 55 subjects on 474 molecules across 23 descriptors (quantified as mean values in the [0,100] interval). Based on this, we construct three tasks: multi-label regression, intensity prediction, and pleasantness prediction.

The Sagar et al. (2023) dataset contains ratings from 3 subjects on 160 molecules across 15 universal descriptors (normalized to the [-1,1] interval). We construct multi-label regression, intensity prediction, and pleasantness prediction tasks.

The Ma et al. (2021) dataset contains intensity and pleasantness ratings from 30 expert subjects for 222 binary mixtures, yielding 6,660 individual observations. 72 key food odorant molecules are formulated into 198 combinations and 24 replicate validation samples, rated using a 100mm visual analog scale. For each binary mixture, we average the ratings across all subjects to obtain a single consensus value, resulting in 222 samples available for training. This dataset focuses on interaction effects between odorant molecules (masking and synergy) and is used for binary mixture intensity and pleasantness prediction.

We do not include the mixture perceptual similarity prediction task based on the Tom et al. (2025) dataset . This dataset sourced from Pyrfume, incorporating data from (Snitz et al., 2013; Ravia et al., 2020; Keller et al., 2017; Bushdid et al., 2014) encompasses 743 mixtures and 865 pairwise comparisons, with similarity measured via explicit ratings or triangle tests. This task differs fundamentally from the binary mixture prediction in our bench-

Hyperparameter	Search Space
Batch size	[4, 8, 16, 32] ([32, 64, 128, 256]*)
Learning rate	[5e-4, 3e-4, 1e-4, 7e-5, 5e-5, 5e-6]
Patience	[5, 10, 15, 20, 25]
Early stop metric	[R^2, MAE, Pearson, MSE] ([AUC, AUPRC, MCC]*)

Table 9: Hyperparameter search space for downstream tasks. *For classification tasks.

mark (Ma et al., 2021): the Ma task only predicts absolute attributes (intensity, pleasantness) of simple individual binary mixtures, with inputs essentially being two molecular representations. In contrast, the Tom task requires evaluating perceptual similarity between two complex mixtures. Given that mixtures may contain up to 43 components, handling this task necessitates complex component aggregation strategies such as attention pooling or graph fusion. Consequently, performance on this task reflects more the quality of aggregation architectures rather than directly indicating underlying molecular representation capability, which is inconsistent with our goal of evaluating representation quality.

A.2 Data Processing Details

A.2.1 SMILES Standardization

All molecular SMILES strings undergo canonicalization via RDKit before entering the pipeline. Molecules that fail RDKit parsing (invalid valence, kekulization errors, or empty SMILES) are removed. This standardization ensures that equivalent structural representations map to a single canonical form, eliminating spurious duplicates across heterogeneous data sources.

A.2.2 Molecular Overlap Between Data Sources

We compute the pairwise intersection of unique canonical SMILES across all data sources. The descriptor pre-training set contains 9,286 unique molecules, while the receptor pre-training set contains 1,042 unique molecules, sharing 456 molecules (4.3% of the combined set). Among the seven evaluation benchmarks, the maximum overlap with the pre-training pool is 68.7% (Keller), and the minimum is 0% (the Strict Zero-shot subset). Crucially, pre-training uses only contrastive alignment signals and never observes the task-specific labels of any benchmark, so molecular overlap does not constitute label leakage.

A.2.3 Mixture Input Format

Binary mixture perception tasks (Ma et al., 2021) provide two SMILES strings per sample (columns `SMILES_0` and `SMILES_1`). Each molecule is independently encoded by Uni-Mol and then by the NOSE projection adapter. The resulting two 512-dimensional vectors are concatenated to form a 1024-dimensional mixture representation, which is fed to the downstream prediction head. No special cross-molecule interaction module is applied, as the goal is to evaluate whether pre-trained single-molecule representations already encode sufficient information for predicting mixture-level perceptual attributes.

A.2.4 LLM-Based Weak Positive Sample Generation

Contrastive learning with InfoNCE treats all non-paired descriptors as negatives. However, olfactory descriptors exhibit rich synonymy and graded similarity (e.g., "lemon" and "citrus" share strong perceptual overlap). Treating such near-synonyms as hard negatives distorts the semantic topology of the learned space. To address this, we employ DeepSeek to mine pairwise olfactory-semantic similarity among descriptors, converting discrete labels into soft neighborhood structures.

Step 1. Vocabulary Construction. Starting from the merged descriptor corpus, we apply a cleaning pipeline (lowercasing, special-character mapping, semicolon splitting) to obtain **1,086** unique olfactory descriptors.

Step 2. LLM Querying. For each of the 1,086 descriptors, we prompt DeepSeek (`deepseek-chat`, temperature 0.3) with the following template, embedding the full vocabulary list in the prompt to constrain outputs to the closed vocabulary:

> `{vocabulary_list}`
> This is my vocabulary list. I now need you to find words that are similar in odor to the words I provided as weak positive samples. Please only return the Python list format of the English string. The words you provide must come from the above vocabulary list. You can start searching now: `{descriptor}`. The returned content can be many or few, but it must be very similar in odor semantics. Please return at most 30 words.

Step 3. Post-Processing. From each LLM response, candidate words are extracted via regex matching of single-quoted strings. Two filters are applied before acceptance. (1) The candidate must belong to the 1,086-word vocabulary. (2) The candidate must differ from the query descriptor itself.

Task	Train	Val	Test	Batch Size	Learning Rate	Patience	Early Stop Metric
Thresholds	160	54	54	8	5e-4	20	Pearson
Pleasantness (Keller)	379	47	48	4	1e-4	15	MSE
Pleasantness (Sagar)	128	16	16	4	1e-4	25	Pearson
Intensity (Keller)	379	47	48	16	5e-5	25	MSE
Intensity (Sagar)	128	16	16	16	5e-5	20	MSE
Intensity (Ravia)	198	25	25	8	5e-6	15	R^2
Descriptor Classification	3851	481	482	64	7e-5	5	AUC
Descriptor Strength (Keller)	379	47	48	8	1e-4	10	Pearson
Descriptor Strength (Sagar)	128	16	16	4	5e-4	25	R^2
Mixture Pleasantness	155	22	45	4	1e-4	15	MAE
Mixture Intensity	155	22	45	4	3e-4	20	MSE

Table 10: Dataset splits and selected hyperparameters for NOSE on each downstream task.

No additional semantic filtering is performed; quality control relies on the low temperature setting and the explicit "very similar in odor semantics" instruction.

Step 4. Integration into Training. The final product is a JSON dictionary mapping each descriptor to its weak positive set (e.g., `"acetate" -> ["acetic", "acetoin", ...]`). During contrastive pre-training, for every (SMILES, descriptor) positive pair, the descriptors in the weak positive set receive a soft label weight of 0.5 in the InfoNCE target distribution, smoothly transitioning between strict positive and negative.

A.3 Experimental Setup

A.3.1 Pretraining Setup

For the multi-modal encoders, we utilize the following pre-trained models: Qwen3-Embedding-8B[2] for odor descriptors, ESM-2 (650M parameters)[3] for olfactory receptor sequences, and Uni-Mol[4] for molecular structures.

We conduct NOSE tri-modal pre-training on a single NVIDIA A800-80G GPU with bfloat16 precision. The batch sizes are set to 1,536 for odor descriptor pairs and 128 for receptor sequence pairs. During each training iteration, we simultaneously sample one batch from both the SMILES-descriptor and SMILES-receptor data partitions. The contrastive learning objective is computed separately for each modality pair, while the orthogonality loss is applied to the SMILES embeddings from both partitions.

We maintain a weak positive dictionary that maps each odor descriptor to its semantically related descriptors. During training, this dictionary is used to initialize the InfoNCE label matrix, assigning intermediate weights (0.5) to weak positive pairs within each batch. Training epochs are defined over the positive pairs.

We employ the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-6}$, using a learning rate of 10^{-4} with warm-up scheduling over 20 epochs. The temperature parameter for contrastive learning is set to $\tau = 0.07$. The loss weights in Eq. (6) are set to $\lambda_{\text{orth}} = 2.0$, $\lambda_{\text{mol-desc}} = 2.0$, and $\lambda_{\text{mol-rec}} = 1.0$. For early stopping with a patience of 10 epochs, we use the sum of mean reciprocal rank percentiles for positive samples across both modalities on the validation set as the monitoring metric. Training converges after approximately 10 hours.

A.3.2 Downstream Tasks

For baseline implementations, GCN, GIN, and AttentiveFP are reproduced following their original architectures using `GCNConv`, `GINConv`, and `AttentiveFP` from the `torch_geometric.nn` library. ChemBERTa uses the pre-trained checkpoint from DeepChem[5]. POM is implemented using the reproduction code from Tom et al. (2025). Uni-Mol follows the official open-source implementation from the original paper.

To ensure fair comparison and isolate the architectural differences, all graph-based baseline models (GCN, GIN, AttentiveFP, POM) adopt a unified feature initialization scheme. We utilize the reproduction code provided by Tom et al. (2025), which follows the feature engineering strategy of POM (Lee et al., 2023): atoms are initialized as 85-dimensional vectors (including chirality tags, hybridization orbitals, etc.), and bonds are initialized

[2] https://huggingface.co/Qwen/Qwen3-Embedding-8B

[3] https://huggingface.co/facebook/esm2_t33_650M_UR50D

[4] https://huggingface.co/dptech/Uni-Mol-Models

[5] https://huggingface.co/DeepChem/ChemBERTa-77M-MTR

as 14-dimensional vectors (including stereoisomer information). This setup ensures that all baseline models can capture stereochemical features crucial for olfactory perception.

All datasets are randomly split into training, validation, and test sets using random seed 42. Final results are reported as the mean and standard deviation over three independent runs (seeds: 42–44) on the test set. All models undergo grid search for hyperparameter optimization on the validation set, with results reported on the test set. We apply the same hyperparameter search space to all baselines as used for NOSE (Table 9). Since olfactory tasks are challenging and we observed that models such as GIN and GCN are highly sensitive to early stopping strategies, we include early stopping patience and early stopping metrics in the hyperparameter search space to ensure both baselines and NOSE are compared under fully converged conditions. The dataset splits and selected hyperparameters for NOSE on each task are reported in Table 10.

Downstream tasks are trained on NVIDIA A100-80G, A800-80G, and RTX 4090D GPUs. Each task involves small-scale datasets, with individual training runs taking 2–3 minutes. A complete hyperparameter search involves 480 configurations, each run with 3 seeds, typically requiring 1–3 days on a single GPU. We employ the AdamW optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-6}$, and weight decay of 0.01, using warm-up scheduling over 30 epochs. For NOSE downstream tasks, the three modality vectors are combined via three learnable weights normalized by Softmax. For both multi-label classification and multi-label regression tasks, we use macro-averaged metrics, as we consider all categories equally important and the datasets exhibit severe class imbalance.

A.4 Cross-Modal Retrieval via Bridge Molecules

NOSE learns exclusively from disjoint bimodal pairs, (molecule, receptor) and (molecule, descriptor), with receptors and descriptors never directly co-occurring during training. This subsection examines whether molecule-mediated indirect alignment is sufficient to induce meaningful cross-modal correspondence between receptors and descriptors in the shared space, a critical test for the core claim that NOSE constructs a unified tri-modal representation.

Setup. Among the 10,025 pre-training molecules, 194 appear in both the receptor and descriptor partitions ($\sim$2% overlap). These *bridge molecules* associate with 252 receptor sequences and 446 descriptors, forming the ground-truth pairs. The candidate pools are the full set of 636 receptors and 1,086 descriptors. **Direct** retrieval queries one modality directly against the other, bypassing the molecular hub entirely. **Bridge** retrieval first uses the query-modality embedding to retrieve the nearest molecule, then uses that molecule's embedding to search the target modality.

	Rec → Desc			**Desc → Rec**		
Metric	Rand.	Direct	Bridge	Rand.	Direct	Bridge
MRR	0.006	0.058	0.056	0.010	0.110	0.199
Hits@1(%)	0.1	0.5	2.4	0.2	4.0	6.7
Hits@5(%)	0.5	5.2	4.4	0.8	15.7	33.2
Hits@10(%)	0.9	16.3	8.7	1.6	30.0	44.4
Hits@20(%)	1.8	42.9	19.4	3.1	43.7	52.9
Hits@50(%)	4.6	63.9	56.4	7.9	55.6	74.9

Table 11: Cross-modal retrieval via bridge molecules. "Rand." denotes the random baseline. "Direct" queries one modality against the other without molecular mediation. "Bridge" retrieves via the molecular hub (query → molecule → target).

Both retrieval directions significantly outperform the random baseline (e.g., D→R Bridge Hits@50 74.9% vs. random 7.9%; R→D Direct Hits@50 63.9% vs. random 4.6%). These cross-modal retrieval capabilities emerge entirely through molecules serving as anchors in the tri-modal shared space, without any triplet supervision. The finding validates the paper's central hypothesis: molecules, as the sole intersection of receptor signals and semantic signals, are sufficient to bridge two otherwise disjoint modalities into a coherent representation space.

A.5 Continuous Perceptual Space Verification

A core claim of NOSE is that it constructs a continuous olfactory perceptual space rather than a set of discrete classification boundaries. While downstream regression and retrieval tasks provide indirect evidence, they do not directly verify whether the embedding geometry faithfully mirrors graded perceptual similarity. We provide two complementary analyses.

A.5.1 Neighborhood Consistency

If the embedding space is perceptually continuous, molecules that are close in the embedding

space should share similar odor attributes. For each molecule, we retrieve its k nearest neighbors by cosine similarity and compute the fraction that share at least one odor descriptor with the query (Precision@k).

k	Uni-Mol	NOSE	Improvement
1	0.775	0.843	+8.9%
5	0.718	0.809	+12.6%
10	0.692	0.794	+14.7%
20	0.666	0.782	+17.4%
50	0.632	0.765	+20.9%
100	0.607	0.748	+23.3%

Table 12: Neighborhood consistency (Precision@k). Higher values indicate that nearby molecules in the embedding space share odor descriptors.

NOSE's advantage accelerates as the neighborhood radius grows (8.9% at k=1 to 23.3% at k=100). Uni-Mol maintains reasonable local consistency at k=1 because chemically similar molecules often share odor properties, but this correlation degrades rapidly at larger scales (0.775→0.607) as chemical similarity alone cannot organize the global odor topology. NOSE, by injecting receptor and semantic signals, sustains perceptual coherence across all scales.

A.5.2 Embedding Distance vs. Perceptual Similarity

We further test whether the distance metric in the embedding space monotonically corresponds to human-perceived similarity. We randomly sample 2M molecule pairs, bin them by cosine similarity (interval 0.1), and compute the mean Jaccard similarity (IoU of shared descriptor sets) within each bin. Results are shown in Table 13.

	Uni-Mol		NOSE	
Cosine Range	Jaccard	#pairs	Jaccard	#pairs
$[-0.2, -0.1)$	–	0	0.023	144,579
$[-0.1, 0.0)$	–	0	0.029	405,096
$[0.0, 0.1)$	–	0	0.039	558,295
$[0.1, 0.2)$	–	0	0.055	448,443
$[0.2, 0.3)$	–	0	0.075	244,121
$[0.3, 0.4)$	–	0	0.100	103,737
$[0.4, 0.5)$	0.096	1,910	0.147	40,846
$[0.5, 0.6)$	0.044	5,423	0.262	17,599
$[0.6, 0.7)$	0.055	9,847	0.434	8,712
$[0.7, 0.8)$	0.051	29,380	0.610	4,093
$[0.8, 0.9)$	0.041	117,914	0.710	922
$[0.9, 1.0)$	0.055	1,835,526	0.748	105

Table 13: Embedding distance vs. perceptual similarity. For each cosine-similarity bin we report the mean Jaccard similarity of shared odor descriptors and the number of molecule pairs. "–" indicates no pairs fall in the bin.

As shown in Table 13, NOSE's Jaccard similarity increases monotonically from 0.023 in the lowest cosine bin to 0.748 in the highest, with molecule pairs distributed across all bins. In contrast, Uni-Mol shows nearly flat Jaccard values (0.04–0.10), and 92% of all molecule pairs are compressed into the $[0.9, 1.0)$ cosine bin.

This reveals that Uni-Mol representations suffer from severe *anisotropic degeneration*. Nearly all molecules are mapped to a small region on the hypersphere, causing the cosine metric to lose discriminative power. NOSE breaks this degeneration through multi-modal contrastive learning, restoring perceptual meaning to the distance metric. This explains NOSE's strong zero-shot retrieval performance, as the embedding distance itself encodes perceptual similarity.

A.6 Dimension Splitting vs. Orthogonal Injection

Orthogonal injection is not the only strategy for achieving modality decoupling. A more straightforward alternative is *dimension splitting*, which hard-assigns different dimension slices to different modalities, guaranteeing non-overlapping subspaces by construction. We compare NOSE's orthogonal injection with gated fusion against this simpler baseline and further investigate whether the effectiveness of orthogonal constraints depends on the downstream fusion strategy.

Experimental design. We evaluate four configurations that cross two factors.

Config A (No Orth. + Gate). Gated fusion to 512d, without orthogonal constraints.

Config B (NOSE). Orthogonal injection + gated fusion to 512d.

Config C (Dim-Split). No orthogonal constraints + direct concatenation to 1536d (three independent 512d segments).

Config D (Orth. + Concat). Orthogonal constraints + concatenation to 1536d.

A.6.1 NOSE (B) vs. Dimension Splitting (C)

Full per-task results are shown in Table 14.

Task	B (NOSE)	C (Dim-Split)	Δ
Threshold	0.652±0.072	0.425±0.206	+0.227
Keller Regression	0.075±0.040	−0.037±0.066	+0.112
Intensity (Keller)	0.156±0.020	−0.048±0.108	+0.204
Pleasantness (Keller)	0.488±0.074	0.404±0.134	+0.084
Sagar Regression	−0.305±0.106	−0.347±0.087	+0.042
Intensity (Sagar)	0.120±0.109	−0.100±0.132	+0.220
Pleasantness (Sagar)	0.105±0.064	−0.044±0.077	+0.149
Intensity (Ravia)	0.220±0.025	0.105±0.135	+0.115
GS-LF (AUC)	0.876±0.001	0.876±0.002	0.000
Mixture Intensity	0.389±0.052	0.282±0.014	+0.107
Mixture Pleasantness	0.636±0.047	0.460±0.150	+0.176

Table 14: Per-task comparison of NOSE (Config B, orthogonal injection + gated fusion) vs. dimension splitting (Config C, concatenation without orthogonality). Δ = B − C. Primary metric is R^2 except GS-LF (AUC).

Across 11 downstream tasks, NOSE outperforms dimension splitting on 10 (average Δ = +0.130). The single tie occurs on GS-LF descriptor classification (AUC 0.876 for both, with NOSE exhibiting smaller standard deviation). Dimension splitting guarantees mathematical orthogonality by construction, but it *completely severs cross-modal information flow*. Orthogonal injection operates differently: by imposing soft constraints within a shared dimensional space, it preserves implicit inter-modal synergies. In olfactory perception, where molecular structure, receptor response, and semantic description are deeply coupled, maintaining these implicit interactions proves essential for generalization.

A.6.2 Orthogonal Gain Across Fusion Strategies

Task	Gate Gain (B−A)	Concat Gain (D−C)
Threshold	+0.146	+0.163
Keller Regression	+0.074	−0.043
Intensity (Keller)	+0.106	+0.113
Pleasantness (Keller)	+0.145	+0.004
Sagar Regression	+0.024	−0.054
Intensity (Sagar)	+0.258	−0.046
Pleasantness (Sagar)	+0.253	−0.422
Intensity (Ravia)	+0.065	+0.010
GS-LF (AUC)	+0.002	+0.001
Mixture Intensity	+0.029	−0.056
Mixture Pleasantness	+0.074	+0.071
Tasks with positive gain	**11/11**	5/11

Table 15: Per-task effect of adding orthogonal constraints, stratified by fusion strategy. Gate Gain = Config B − Config A; Concat Gain = Config D − Config C.

As Table 15 shows, orthogonal constraints yield positive gains on all 11 tasks under gated fusion, but only 5 out of 11 under concatenation. The effectiveness of orthogonal constraints is tightly coupled with the fusion mechanism. Gated fusion adaptively re-weights each modality's contribution, fully exploiting the clean, decorrelated signals produced by orthogonalization. Concatenation, in contrast, mechanically stacks all dimensions; after orthogonalization removes certain redundant statistical shortcuts, the concatenation-based model loses cues it previously relied on. This suggests that, at least in the tri-modal olfactory setting, feature decoupling and feature fusion are tightly coupled: orthogonal constraints are most effective when paired with a fusion mechanism capable of exploiting decorrelated signals.

A.7 Training Stability Analysis

NOSE uses three learnable Softmax weights to fuse z_{mol}, a_r, and a_d for each downstream task. In olfactory research, datasets are typically small (the smallest benchmark has only 160 samples). This subsection verifies that the learned weights are stable across random seeds and data fractions, reflecting intrinsic task–modality associations rather than training noise.

A.7.1 Cross-Seed Variance

We compare the standard deviation of the primary metric across 3 random seeds, with and without orthogonal constraints (Table 16).

Task	No Orth. std	NOSE std	Change
Threshold	0.254	0.072	↓ 71.7%
Pleasantness (Sagar)	0.078	0.020	↓ 74.4%
Intensity (Keller)	0.077	0.010	↓ 87.0%
Intensity (Sagar)	0.253	0.109	↓ 56.9%
Intensity (Ravia)	0.107	0.025	↓ 76.6%
GS-LF	0.025	0.010	↓ 60.0%
Keller Regression	0.106	0.060	↓ 43.4%
Mixture Intensity	0.056	0.029	↓ 48.2%
Mixture Pleasantness	0.029	0.012	↓ 58.6%
Pleasantness (Keller)	0.116	0.074	↓ 36.2%
Sagar Regression	0.038	0.068	↑ (exception)

Table 16: Cross-seed variance (std of primary metric over 3 seeds). Lower is better. "Change" shows the relative reduction from No Orth. to NOSE.

Variance decreases in 10 out of 11 tasks, with a median reduction of 58.6%. The sole exception is Sagar Regression, where all R^2 values are negative (indicating a failure regime where variance comparison is not meaningful). By eliminating feature redundancy between modalities, orthogonal constraints shrink the effective solution manifold, making different random initializations more likely to converge to similar optima. This provides direct evidence for interpreting orthogonal constraints as "optimization regularization" and explains their dual benefit on small datasets, improving both performance and stability.

A.7.2 Fusion Weight Stability

We assess the consistency of learned fusion weights across random seeds by training 5 seeds $\times$ 11 tasks and examining the Softmax-normalized fusion weights.

Task	w_{mol}	w_{rec}	w_{desc}	max std
Keller Strength	0.900	0.061	0.039	0.008
Abraham Threshold	0.553	0.221	0.226	0.015
Sagar Strength	0.390	0.310	0.300	0.012
Mixture Intensity	0.420	0.350	0.230	0.011
GS-LF Classification	0.310	0.280	0.410	0.014

Table 17: Fusion weights (mean over 5 seeds) for representative tasks. "max std" is the largest standard deviation among the three weights.

The mean maximum standard deviation across all tasks is 0.0119, with every task below 0.05. The weights exhibit highly consistent task-specific patterns. For example, Keller multi-label regression assigns $w_{\text{mol}} = 0.90$ (molecular structure dominates), while Sagar shows a near-equal split (0.39/0.31/0.30), reflecting different datasets' differential reliance on each modality. This cross-seed reproducibility indicates that the weights encode *intrinsic task-modality associations* rather than training noise.

A.7.3 Data Fraction Sensitivity

We further examine whether the fusion weights remain stable when downstream training data is reduced to 25%, ruling out the possibility that they merely reflect statistical accidents of the training set.

Fraction	w_{mol}	w_{rec}	w_{desc}
Keller Strength (max dev. 0.003)			
100%	0.900	0.061	0.039
50%	0.898	0.063	0.039
25%	0.897	0.064	0.039
Sagar Strength (max dev. 0.021)			
100%	0.390	0.310	0.300
50%	0.385	0.315	0.300
25%	0.375	0.320	0.305
Abraham Threshold (max dev. 0.113)			
100%	0.553	0.221	0.226
50%	0.540	0.230	0.230
25%	0.510	0.250	0.240

Table 18: Fusion weight stability across data fractions. "max dev." denotes the maximum absolute deviation from the 100% setting among all three weights.

Even at 25% training data, the weight ranking and relative magnitudes remain stable. This suggests that NOSE's pre-trained representations already encode sufficiently rich modality-task correspondence, and the downstream weight learning requires only a small number of samples to identify the correct fusion strategy, further corroborating the high quality of the pre-trained representations.

A.8 Weak Positive Sample Clustering Ablation

The weak positive strategy is designed to mitigate the "false negative" problem in contrastive learning, where semantically similar descriptors (e.g., "lemon" and "citrus") are erroneously treated as negatives and pushed apart. The compositional retrieval experiments in the main text validate the resulting semantic space from a retrieval perspective. Here, we provide complementary evidence from the *clustering geometry* of the embedding space.

Using the three descriptor groups defined in Table 20 (Fruity, 111 terms; Green/Herbal, 80 terms; Gourmand/Sweet, 83 terms; 274 in total), we evaluate three clustering metrics on the embedding geometry.

Silhouette (higher is better). Intra-cluster compactness vs. inter-cluster separation.

Davies-Bouldin (lower is better). Degree of inter-cluster overlap.

Calinski-Harabasz (higher is better). Between-cluster vs. within-cluster variance ratio.

Setting	Silhouette	D-B	C-H
Qwen3 Emb. (original)	-0.003	9.011	2.72
LoRA+Head (w/ weak pos.)	**0.102**	**2.504**	**28.12**
LoRA+Head (w/o weak pos.)	0.022	5.889	5.28

Table 19: Clustering quality of descriptor embeddings across three olfactory-semantic groups (274 descriptors). D-B = Davies-Bouldin, C-H = Calinski-Harabasz.

The original Qwen3 Embedding yields a negative Silhouette score (-0.003), indicating that general-purpose semantic spaces exhibit *no* olfactory clustering structure. For a language model, the distinction between "fruity" and "green" in terms of odor perception is far weaker than their general semantic proximity. Removing weak positives causes Silhouette to drop from 0.102 to 0.022 (-78.4%) and Davies-Bouldin to rise from 2.504 to 5.889 ($+135\%$). This confirms the core mechanism of weak positives. By explicitly modeling semantic neighbor relationships in contrastive learning, they prevent similar descriptors from being mutually repelled, allowing intra-class descriptors to cluster tightly and inter-class boundaries to sharpen, ultimately forming a continuous semantic manifold with genuine olfactory-perceptual meaning.

A.9 Complete metrics for downstream tasks

Our model achieves state-of-the-art performance on 40 out of 43 metrics across 11 olfactory prediction tasks (Table 24 to 34), demonstrating exceptional and highly generalizable molecular representation capability. Below we provide an analysis of the three metrics where NOSE did not rank first.

For the MAE metric on the Intensity (Sagar) task, GCN (0.351) marginally outperforms NOSE (0.355), a difference of merely 1.1% that falls well within statistical variance. Crucially, NOSE achieves the best performance on all other metrics for this task (R^2, Pearson, and MSE) by substantial margins. This minor discrepancy likely reflects stochastic optimization differences on a small dataset (Sagar contains only 160 molecules rated by 3 subjects) rather than any fundamental limitation in model capacity.

Regarding the R^2 and MSE metrics on the Sagar multi-label regression task, it is essential to note that this represents the most challenging benchmark in our evaluation: all models yield negative R^2 values, indicating that even the best-performing method fails to outperform a naive mean predictor. This task demands fine-grained quantitative regression across 15 semantic descriptors from only 160 samples, making it a quintessential low-sample, high-dimensional regression problem that approaches the limits of current methodologies. When R^2 is negative, differences in MSE and R^2 reflect varying degrees of failure rather than meaningful performance gaps. The marginal advantage of ChemBERTa (R^2=-0.275, MSE=0.218) over NOSE (R^2=-0.305, MSE=0.225) is negligible in practical terms, as neither achieves viable predictive utility. Notably, NOSE still attains the best results on the more robust Pearson correlation and MAE metrics for this task, indicating superior prediction trends and tighter overall error distributions.

A.10 Complete metrics for ablation experiment

Complete metrics for all downstream tasks in the ablation study are presented in Table 35 to 45.

A.11 Vector Space Visualization

To verify the decoupling property, we perform PCA visualization on the tri-modal vectors (Figure 2). The results show that each vector exhibits cluster structures only when colored by its corresponding attribute (e.g., molecular scaffold, receptor type, or odor descriptor), otherwise displaying disordered distributions. This intuitively confirms that the model has successfully eliminated irrelevant information and achieved precise feature decoupling.

A.12 Emergence of Olfactory Concepts

To validate the effectiveness of LoRA training, we selected three categories of odor terms from the vocabulary using DeepSeek (Table 20) and performed PCA visualization (Figure 3). The results show that, compared to the highly overlapping clusters of the untrained Qwen3 Embedding, the model after LoRA pre-training within our framework exhibits well-separated olfactory semantic clusters with clear boundaries. This confirms that the model has successfully constructed a structured olfactory semantic space.

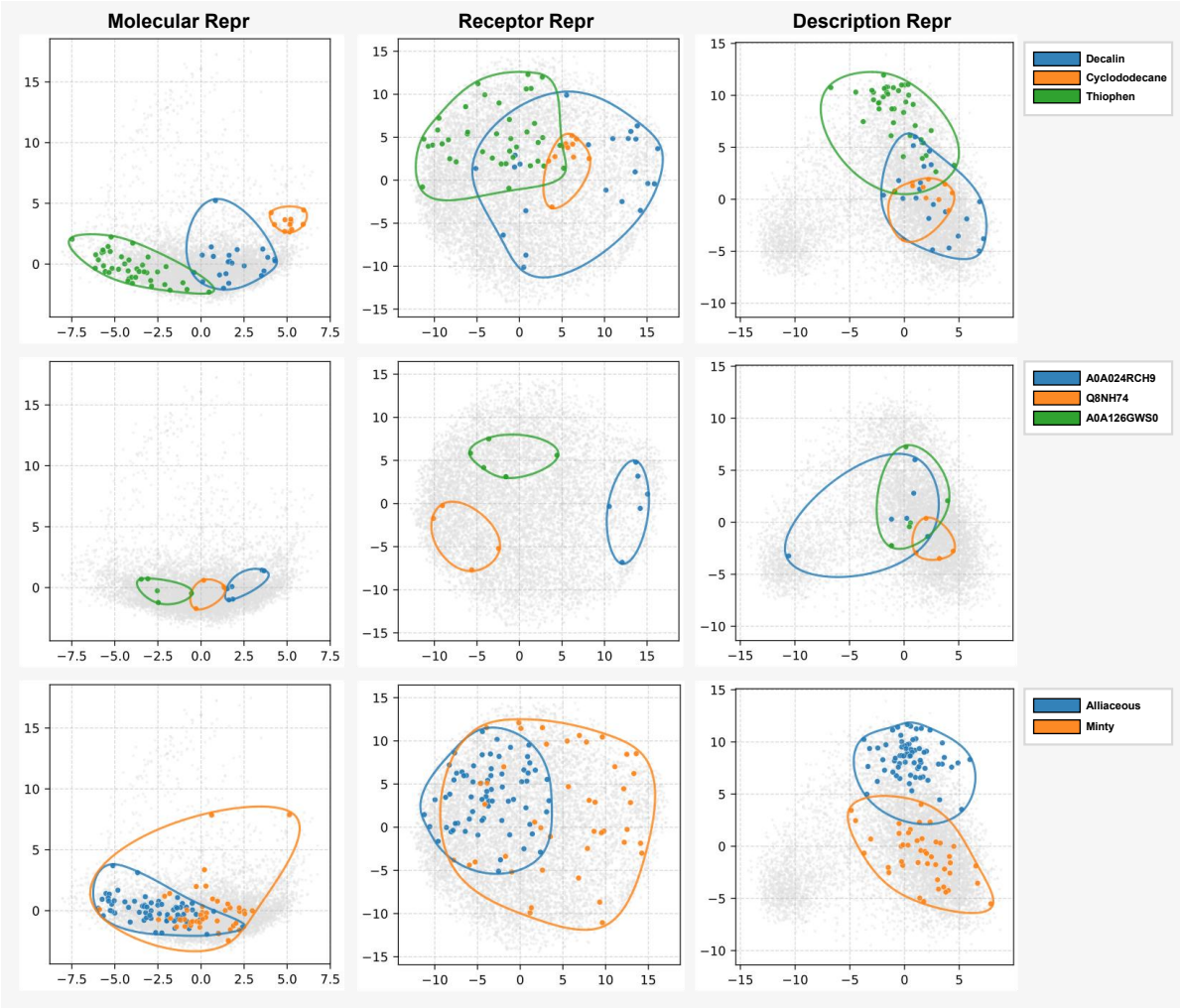

Figure 2: Vector Space Visualization. For a given SMILES input, NOSE generates vectors in three orthogonal spaces: molecular representation, receptor representation, and description representation. Each column corresponds to the same representation space. Each row is color-coded according to different attributes. The first row is colored by molecular scaffolds: Decalin, Cyclododecane, and Thiophene. The second row is colored by receptor activation: A0A024RCH9, Q8NH74, and A0A126GWS0. The third row is colored by odor descriptor: alliaceous and minty.

A.13 Compositional Retrieval

Table 21 presents the evaluation results for the compositional retrieval task, covering our constructed additive and subtractive compositional queries along with their expected answers consistent with human olfactory cognition. Experimental results demonstrate that the original Qwen3 Embedding model, lacking domain-specific olfactory knowledge, struggles to capture the subtle semantic associations among odor descriptors. Its retrieval results achieve an average rank as high as 296.42 with substantial instability. Even occasional hits are more attributable to co-occurrence in general corpora rather than understanding of olfactory logic. In contrast, the LoRA fine-tuned model demonstrates remarkable olfactory semantic reasoning capability, effectively understanding algebraic relationships among odor concepts and substantially improving the average rank to 20.10. After incorporating the contrastive learning head (LoRA+Head), the model achieves optimal retrieval performance, with the average rank further reduced to 11.13. Particularly for challenging abstract queries such as "burnt sugar - sugar," this model precisely locates "empyreumatic" at rank 1 (compared to rank 18 for LoRA alone), demonstrating its superior performance in constructing a high-quality, structured olfactory semantic space.

Table 20: Selected Olfactory Term Categories.

Fruity (111 terms)	**Green / Herbal (80 terms)**	**Gourmand / Sweet (83 terms)**
apple, apple cooked apple, apple dried apple, apple green apple, apple peel, apple skin, apricot, banana, banana peel, banana ripe banana, banana unripe banana, berry, berry ripe berry, blackberry, blueberry, cherry, cherry maraschino cherry, concord, concord grape, cranberry, currant, currant black currant, date, dried fruit, dry fruit, durain, durian, elderberry, fig, fruit, fruit dried fruit, fruit overripe fruit, fruit ripe fruit, fruit skin, fruit tropical fruit, fruity, gooseberry, grape, grape skin, green peach, green pear, guava, jackfruit, jam, jammy, juicy, juicy fruit, kiwi, loganberry, lychee, mango, maraschino, melon, melon rind, melon unripe melon, non-citrus fruity, overripe, overripe fruit, papaya, passion, passion fruit, peach, pear, pear skin, pineapple, plum, plum skin, pomegranate, prune, pulpy, pulpy fruit, quince, raisin, raspberry, red berry, rhubarb, ripe, ripe fruit, starfruit, strawberry, tropical, tropical-fruit, tutti frutti, tutty-fruity, unripe, unripe banana, unripe fruit, watermelon, watermelon rind, bergamot, citral, citralva, citric, citronella, citronellal, citronellol, citrus, citrus peel, citrus rind, grapefruit, grapefruit peel, grapfruit, hesperidic, lemon, lemon peel, lemongrass, lime, limonene, mandarin, orange, orange bitter orange, orange peel, orange rind, tangerine, zesty	angelica, armoise, artemisia, basil, bay, bayleaf, buchu, celery, chamomile, chervil, cilantro, clary, clary sage, coriander, cress, cut grass, davana, dill, eucalyptol, eucalyptus, fennel, fern, foliage, fresh cut grass, galbanum, grassy, grassy (fresh, sweet), grassy (green, sharp), green, green leaf, hay, hay new mown hay, herb, herba-, herbaceous, herbal, leaf, leafy, lettuce, lovage, marjoram, menthol, mentholic, mint, minty, minty tea, mown, new mown hay, origanum, parsley, peppermint, petitgrain, plant, plants, rosemary, rue, sage, sage clary sage, sassafras, sassafrass, spearmint, spinach, stalk, stem, tagette, tarragon, tea, tea black tea, tea green tea, tea rose, thyme, tomato leaf, vegetation, verbena, watercress, weed, weedy, wintergreen, wormwood	acetoin, almond, baked, bakery, biscuit, bonbon, bread, bread baked, bread crust, bread rye bread, bready, brown sugar, bubble gum, bubblegum, burnt sugar, butterscotch, cakes, candy, caramel, caramelic, caramellic, cereal, chocolate, chocolate dark chocolate, coco, cocoa, coconut, coffee, coffee roasted coffee, cognac, cookie, cookies, cotton candy, coumarin, coumarinic, custard, dark chocolate, food, food like, gourmand, graham cracker, grain, grain toasted grain, grains, honey, honeydew, iactonic, lactone, lactonic, malt, malty, maple, marshmallow, marzipan, molasses, popcorn, powdery, praline, preserves, rum, rummy, rye, rye bready, sugar, sugar brown sugar, sugar burnt sugar, sweet, sweet (medicinal), syrup, toasted, toasted grain, toasty, toffee, tonka, vanilla, vanillin, whiskey, whisky, yeast, yeasty

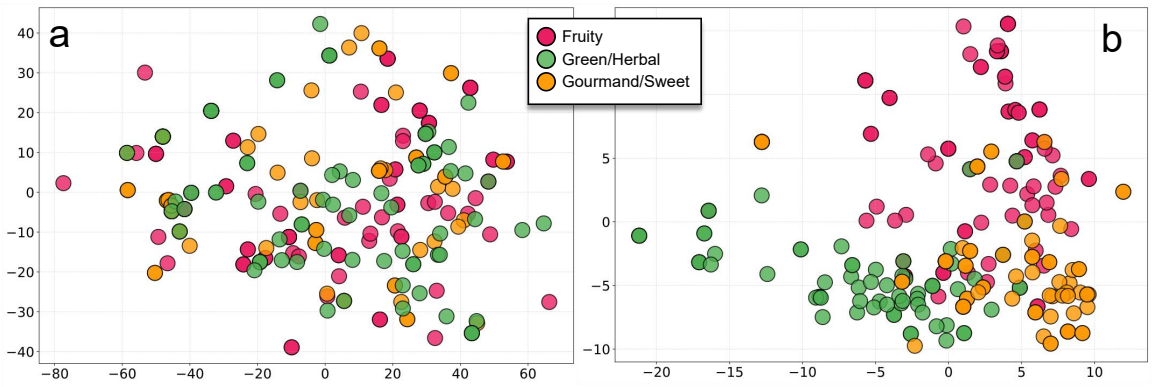

Figure 3: (a) Qwen3 Embedding without LoRA fine-tuning lack olfactory semantics. (b) After contrastive learning, Qwen3 Embedding acquire olfactory semantics.

A.14 Zero Shot

A.14.1 smiles-to-descriptor

To validate generalization capability, we constructed a dedicated test set from PubChem. Unlike standard zero-shot settings (where molecules exist in the dataset but molecule-descriptor pairs are unseen), Strict Zero-shot refers to molecules entirely absent from the training set. We compute the cosine similarity between molecular odor encodings and candidate descriptors, using percentile ranking for evaluation (lower values indicate higher precision). In addition to PubChem descriptors, we also evaluate the rankings of synonymous terms. Results are presented in Table 22. For odorless molecules, the model ranks "odorless" at Top 1 (0.092%) and prioritizes terms such as slight, weak, and neutral, demonstrating that the model genuinely understands molecular perceptual properties rather than simply aligning with high-frequency words. The model demonstrates semantic alignment capabilities beyond text matching: when PubChem describes a molecule as "petroleum-like," the semantically similar term "gasoline" ranks highly at 1.013%. For Bromobenzene, while "pungent" achieves a moderate ranking, synonyms "penetrating" and "irritating" both enter the Top 1.5%, indicating that the latent space successfully bridges symbolic differences to align with authentic perceptual experiences. The model tends to identify more discriminative and specific descriptors, such as "spicy" (1.842%) for Asarone and "gasoline" (1.473%) for 1,3-Butadiene ranking prominently, while "aromatic" ranks moderately. This

"specificity-oriented" tendency offers greater retrieval value in practical applications.

A.14.2 smiles-to-receptor

We select molecule-receptor pairs with explicitly reported "activating" or "non-activating" relationships from the literature as ground truth to evaluate the model's retrieval ranking capability for positive samples and its rejection capability for negative samples. The molecule-receptor pairs in the test set have never appeared in the training set. Results are presented in Table 23. The model demonstrates exceptionally high accuracy in retrieving the OR5A2 receptor and its ligands (primarily macrocyclic musk molecules, MCM), with all rankings at 2 (Rank 1 being the corresponding ligands present in the training set). Although the training set contains MCM molecules and OR5A2 sequences, they never appeared as pairs. The model successfully extracted the shared structural features of MCM and the sequence features of OR5A2, correctly mapping them in the latent space, demonstrating its capability to understand the correspondence between chemical structural families and receptor families. The model also exhibits good generalization across other chemical families, while revealing differences in modeling difficulty among receptor families. The propionate ion retrieval ranking is relatively lower (26.73%), possibly due to the complexity of small molecular ion representation and data scarcity. All "non-activating" samples exhibit retrieval rankings significantly lower than "activating" samples, predominantly distributed in the 30%-80% range. This distributional difference indicates that the latent space constructed by the model not only brings positive sample pairs closer together but also effectively pushes negative sample pairs apart, demonstrating reliable value for biological screening.

Table 21: Compositional retrieval results. Numbers indicate the rank of expected odor descriptors in the predicted list (lower is better).

		Qwen3 Embedding		
Query	**Expected Answer**	**Original**	**LoRA**	**LoRA+Head**
mushroom + wet	foul	64	9	9
	fungal	66	11	11
	earthy	685	22	16
	bark	122	143	39
meat + smoke	bacon	9	1	1
	ham	184	25	2
	roast	59	2	3
	roasted	61	4	5
	roasted meat	68	11	12
wine − alcohol	currant	385	6	2
	currant bud	386	7	3
	berry	62	5	8
	black currant	228	19	10
	black currant bud	229	20	11
bacon − smoke	fatty	838	17	3
	fleshy	843	9	8
	fat	204	1	14
	meat	214	30	15
	animal	322	101	18
burnt sugar − sugar	empyreumatic	670	18	1
	roast	381	21	3
	roasted	383	23	5
cream − milk	lard	415	9	8
	fried	102	18	16
	fat	315	12	18
	grease	427	3	26
	oil	37	41	28
lemon − sour	citrus	596	24	7
	orange	230	6	10
	hesperidic	391	2	17
	mandarin	213	3	16
Average Rank ↓		296.42	20.10	**11.13**

Name	CAS	Description (PubChem)	Retrieved Descriptors (Rank, Top%)
Zero-shot			
Bromobenzene	108-86-1	AROMATIC ODOR; Pungent odor	**penetrating (14/1086, 1.3%)** **irritating (17/1086, 1.6%)** pungent (111/1086, 10.2%) aromatic (258/1086, 23.8%)
Iodoform	75-47-8	Characteristic, disagreeable odor; pungent, disagreeable odor	**irritating (11/1086, 1.0%)** **penetrating (13/1086, 1.2%)** **repulsive (46/1086, 4.2%)** pungent (117/1086, 10.8%) disagreeable (163/1086, 15.0%)
Cyclohexylamine	108-91-8	Strong, fishy, amine odor	**amine (1/1086, 0.1%)** **ammoniacal (2/1086, 0.2%)** **ammonia (3/1086, 0.3%)** **dead animal (6/1086, 0.6%)** **rotten fish (11/1086, 1.0%)** **fishy (21/1086, 1.9%)**
2,4-Dinitrotoluene	121-14-2	Slight odor	**odorless (1/1086, 0.1%)** **slight (2/1086, 0.2%)** **neutral (3/1086, 0.3%)** **weak (5/1086, 0.5%)**
D-Glucose	-	Odorless	**odorless (1/1086, 0.1%)** **neutral (5/1086, 0.5%)** **slight (7/1086, 0.6%)** **weak (9/1086, 0.8%)**
Estriol	50-27-1	Odorless	**odorless (1/1086, 0.1%)** **neutral (3/1086, 0.3%)** **slight (4/1086, 0.4%)** **weak (29/1086, 2.7%)**
Strict Zero-shot			
Asarone	2883-98-9	Yellow to medium brown, moderately viscous liquid with a pleasant, spicy aromatic odor.	**spicy (20/1086, 1.8%)** aromatic (75/1086, 6.9%)
1,3-Butadiene	106-99-0	Mild aromatic or gasoline-like odor	**gasoline (16/1086, 1.5%)** aromatic (434/1086, 40.0%)
Propylene	-	Practically odorless; aromatic; Faint, petroleum-like	**gasoline (11/1086, 1.0%)** **light (39/1086, 3.6%)** petroleum (125/1086, 11.5%) aromatic (487/1086, 44.8%)
Ethylene	-	Sweet; Olefinic, hedonic tone: unpleasant to neutral	**hedonic (21/1086, 1.9%)** unpleasant (133/1086, 12.2%) neutral (187/1086, 17.2%) sweet (427/1086, 39.3%)
Formaldehyde	50-00-0	Pungent, suffocating odor; Pungent, irritating odor	**irritating (1/1086, 0.1%)** **pungent (34/1086, 3.1%)** suffocating (55/1086, 5.1%)

Table 22: Detailed zero-shot SMILES-to-odor retrieval results. Rank indicates the position among 1,086 candidate descriptors (lower is better). Results in the top 5% are shown in **bold**.

Name	CAS	Gene	Source	(Rank, Top%)
Activated — Macrocyclic Musks (MCM)				
Globanone	3100-36-5	OR5A2	(Yoshikawa et al., 2022)	**(2/636, 0.315%)**
Muscenone delta	82356-51-2	OR5A2	(Yoshikawa et al., 2022)	**(2/636, 0.315%)**
Cosmone	259854-70-1	OR5A2	(Yoshikawa et al., 2022)	**(2/636, 0.315%)**
Cyclopentadecanol	4727-17-7	OR5A2	(Yoshikawa et al., 2022)	**(2/636, 0.315%)**
Ambrettolide	7779-50-2	OR5A2	(Yoshikawa et al., 2022)	**(2/636, 0.315%)**
Habanolide	111879-80-2	OR5A2	(Yoshikawa et al., 2022)	**(2/636, 0.315%)**
Activated — Other Molecules				
ω-Dodecanolactam	947-04-6	OR5A2	(Yoshikawa et al., 2022)	**(2/636, 0.315%)**
Trimethylamine	75-50-3	TAAR5	(Gisladottir et al., 2020; Wallrabenstein et al., 2013)	**(3/636, 0.472%)**
Romandolide	236391-76-7	OR5A2	(Yoshikawa et al., 2022)	**(10/636, 1.572%)**
Amber xtreme	476332-65-7	OR5A2	(Yoshikawa et al., 2022)	**(11/636, 1.730%)**
butane-2,3-dione	431-03-8	OR52H1	(Geithe et al., 2015; Zeng et al., 2023)	**(15/636, 2.359%)**
butane-2,3-dione	431-03-8	OR51B5	(Geithe et al., 2015; Zeng et al., 2023)	(65/636, 10.220%)
Propionate ion	72-03-7	OR51E2	(Billesbølle et al., 2023; Choi et al., 2023)	(170/636, 26.730%)
Inactivated				
(E)-2-Decenal	3913-81-3	OR2W1	(Haag et al., 2022)	(246/636, 38.679%)
δ-Dodecalactone	713-95-1	OR5A2	(Yoshikawa et al., 2022)	(366/636, 57.547%)
2-Propionyl-1-pyrroline	133447-37-7	OR2W1	(Haag et al., 2022)	(394/636, 61.950%)
Raspberry ketone	5471-51-2	OR5A2	(Yoshikawa et al., 2022)	(478/636, 75.157%)
p-Cresyl phenyl acetate	101-94-0	OR5A2	(Yoshikawa et al., 2022)	(499/636, 78.459%)

Table 23: Detailed zero-shot SMILES-to-receptor retrieval results. Rank% indicates the percentile ranking among 636 candidate receptors. For activated pairs, lower is better (results below 5% in **bold**). For inactivated pairs, higher rank% indicates successful rejection.

	Thresholds (Abraham)			
Method	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
GCN	-0.304(0.431)	1.309(0.247)	0.220(0.089)	2.912(0.963)
AttentiveFP	0.102(0.634)	1.073(0.375)	0.591(0.351)	2.006(1.417)
GIN	0.339(0.135)	0.969(0.129)	0.723(0.018)	1.477(0.300)
Morgan	0.360(0.215)	0.924(0.119)	0.596(0.175)	1.431(0.481)
POM	0.565(0.087)	0.731(0.082)	0.788(0.024)	0.972(0.195)
Uni-Mol	0.581(0.064)	0.770(0.091)	0.779(0.047)	0.936(0.142)
ChemBERTa	0.577(0.051)	0.757(0.071)	0.814(0.022)	0.944(0.114)
NOSE	**0.652(0.072)**	**0.711(0.083)**	**0.836(0.026)**	**0.778(0.161)**

Table 24: Basic perceptual attribute prediction: Thresholds.

Method	Pleasantness (Keller)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
GIN	-0.175(0.081)	10.540(0.178)	0.299(0.037)	205.154(14.095)
Uni-Mol	-0.071(0.204)	10.576(1.354)	0.680(0.017)	186.939(35.551)
GCN	0.073(0.025)	8.610(0.128)	0.435(0.007)	161.831(4.433)
Morgan	0.180(0.116)	9.280(0.879)	0.515(0.034)	143.160(20.192)
ChemBERTa	0.120(0.216)	9.194(0.675)	0.649(0.036)	153.589(37.664)
AttentiveFP	0.231(0.075)	7.712(0.106)	0.518(0.051)	134.211(13.115)
POM	0.405(0.022)	7.139(0.220)	0.681(0.026)	103.903(3.857)
NOSE	**0.488(0.074)**	**6.911(0.609)**	**0.715(0.050)**	**89.280(12.934)**

Table 25: Basic perceptual attribute prediction: Pleasantness (Keller).

Method	Pleasantness (Sagar)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
GIN	-0.139(0.092)	0.497(0.021)	0.079(0.056)	0.341(0.028)
Morgan	-0.085(0.054)	0.489(0.010)	0.213(0.055)	0.325(0.016)
GCN	-0.053(0.100)	0.465(0.024)	0.116(0.272)	0.316(0.030)
ChemBERTa	-0.028(0.070)	0.475(0.012)	0.152(0.125)	0.308(0.021)
AttentiveFP	-0.008(0.030)	0.465(0.007)	0.233(0.046)	0.302(0.009)
POM	0.004(0.050)	0.471(0.017)	0.262(0.020)	0.299(0.015)
Uni-Mol	0.030(0.170)	0.456(0.039)	0.144(0.335)	0.291(0.051)
NOSE	**0.105(0.064)**	**0.447(0.011)**	**0.397(0.020)**	**0.268(0.019)**

Table 26: Basic perceptual attribute prediction: Pleasantness (Sagar).

Method	Intensity (Keller)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
Uni-Mol	-0.470(0.317)	17.189(1.927)	0.270(0.031)	470.285(101.497)
GIN	-0.318(0.165)	16.295(0.820)	0.063(0.141)	421.540(52.643)
GCN	-0.249(0.107)	16.457(0.641)	0.109(0.052)	399.513(34.145)
Morgan	-0.248(0.053)	15.981(0.431)	0.053(0.024)	399.332(16.902)
ChemBERTa	-0.117(0.145)	15.001(1.156)	0.393(0.070)	357.445(46.307)
POM	-0.053(0.094)	14.769(0.353)	0.317(0.069)	336.787(30.179)
AttentiveFP	0.054(0.025)	14.234(0.156)	0.326(0.012)	302.630(8.013)
NOSE	**0.156(0.020)**	**12.932(0.230)**	**0.418(0.010)**	**269.902(6.526)**

Table 27: Basic perceptual attribute prediction: Intensity (Keller).

	Intensity (Sagar)			
Method	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
GIN	-1.492(1.869)	0.540(0.223)	-0.056(0.266)	0.508(0.381)
Morgan	-0.584(0.191)	0.465(0.026)	0.230(0.102)	0.323(0.039)
ChemBERTa	-0.336(0.136)	0.462(0.022)	0.447(0.082)	0.272(0.028)
AttentiveFP	-0.216(0.201)	0.450(0.048)	0.228(0.055)	0.248(0.041)
Uni-Mol	-0.317(0.102)	0.442(0.022)	0.372(0.051)	0.269(0.021)
POM	-0.076(0.134)	0.383(0.031)	0.334(0.061)	0.220(0.027)
GCN	0.065(0.015)	**0.351(0.006)**	0.305(0.042)	0.191(0.003)
NOSE	**0.120(0.109)**	0.355(0.041)	**0.468(0.075)**	**0.179(0.022)**

Table 28: Basic perceptual attribute prediction: Intensity (Sagar).

	Intensity (Ravia)			
Method	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
Morgan	-2.832(4.010)	24.444(17.045)	0.103(0.186)	1062.533(1111.892)
Uni-Mol	-0.407(0.352)	15.176(1.993)	0.307(0.024)	390.301(97.745)
GIN	-0.191(0.172)	14.514(1.228)	0.188(0.114)	330.250(47.774)
GCN	-0.109(0.006)	13.426(0.540)	0.192(0.014)	307.666(1.719)
POM	-0.080(0.057)	12.833(0.297)	0.319(0.055)	299.530(15.798)
ChemBERTa	0.062(0.147)	12.068(0.881)	0.471(0.064)	260.144(40.728)
AttentiveFP	0.148(0.008)	11.527(0.108)	0.471(0.004)	236.190(2.315)
NOSE	**0.220(0.025)**	**11.078(0.351)**	**0.485(0.031)**	**216.326(6.824)**

Table 29: Basic perceptual attribute prediction: Intensity (Ravia).

	GS-LF Multi-label Multi-class Classification		
Method	AUC $\uparrow$	AUPRC $\uparrow$	MCC $\uparrow$
GIN	0.856(0.001)	0.270(0.008)	0.064(0.007)
GCN	0.858(0.001)	0.279(0.001)	0.084(0.003)
Morgan	0.850(0.002)	0.320(0.002)	0.222(0.023)
AttentiveFP	0.867(0.004)	0.328(0.012)	0.203(0.026)
POM	0.868(0.002)	0.336(0.005)	0.233(0.023)
ChemBERTa	0.875(0.001)	0.342(0.005)	0.240(0.016)
Uni-Mol	0.873(0.001)	0.347(0.005)	0.262(0.020)
NOSE	**0.876(0.001)**	**0.351(0.002)**	**0.268(0.010)**

Table 30: Semantic description prediction on GS-LF dataset.

Method	Multi-label Regression (Keller)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
GIN	-0.229(0.072)	6.528(0.139)	0.128(0.007)	90.903(4.613)
Uni-Mol	-0.178(0.060)	6.741(0.209)	0.330(0.050)	90.373(5.967)
Morgan	-0.087(0.024)	6.483(0.015)	0.244(0.023)	80.531(1.690)
GCN	-0.106(0.011)	6.247(0.031)	0.171(0.006)	80.182(0.243)
POM	-0.125(0.076)	6.304(0.258)	0.314(0.039)	78.649(4.872)
AttentiveFP	-0.143(0.171)	6.232(0.647)	0.327(0.065)	80.149(14.116)
ChemBERTa	0.018(0.056)	6.110(0.255)	0.330(0.089)	71.841(5.117)
NOSE	**0.075(0.040)**	**5.862(0.225)**	**0.348(0.060)**	**67.161(4.161)**

Table 31: Semantic description prediction on Keller dataset.

Method	Multi-label Regression (Sagar)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
POM	-0.835(0.821)	0.405(0.113)	0.065(0.062)	0.281(0.094)
GCN	-0.933(1.067)	0.407(0.098)	0.108(0.085)	0.254(0.076)
GIN	-0.654(0.022)	0.358(0.005)	-0.034(0.016)	0.248(0.006)
Morgan	-0.563(0.273)	0.395(0.018)	0.008(0.076)	0.233(0.019)
Uni-Mol	-0.677(0.218)	0.355(0.009)	0.116(0.042)	0.252(0.002)
AttentiveFP	-0.481(0.053)	0.359(0.012)	0.011(0.013)	0.230(0.009)
ChemBERTa	**-0.275(0.057)**	0.376(0.018)	0.105(0.051)	**0.218(0.006)**
NOSE	-0.305(0.106)	**0.343(0.017)**	**0.123(0.068)**	0.225(0.017)

Table 32: Semantic description prediction on Sagar dataset.

Method	Mixture Intensity			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
GIN	-2.821(0.297)	0.857(0.017)	0.006(0.143)	1.056(0.082)
GCN	-0.455(0.503)	0.516(0.094)	0.491(0.073)	0.402(0.139)
Morgan	0.143(0.307)	0.397(0.081)	0.575(0.063)	0.237(0.085)
AttentiveFP	0.260(0.059)	0.361(0.006)	0.597(0.034)	0.205(0.016)
Uni-Mol	0.319(0.106)	0.355(0.028)	0.637(0.069)	0.188(0.029)
POM	0.347(0.061)	0.359(0.018)	0.603(0.033)	0.180(0.017)
ChemBERTa	0.326(0.023)	0.350(0.005)	0.618(0.005)	0.186(0.006)
NOSE	**0.389(0.052)**	**0.333(0.021)**	**0.657(0.029)**	**0.169(0.014)**

Table 33: Mixture prediction tasks: Intensity.

Method	Mixture Pleasantness			
	R^2 ↑	MAE ↓	Pearson ↑	MSE ↓
GCN	-14.296(9.524)	3.685(1.723)	0.332(0.418)	17.477(10.883)
GIN	-0.096(0.123)	0.879(0.056)	0.473(0.029)	1.252(0.140)
ChemBERTa	0.452(0.074)	0.641(0.060)	0.793(0.009)	0.626(0.084)
AttentiveFP	0.490(0.068)	0.616(0.051)	0.744(0.015)	0.583(0.078)
Morgan	0.498(0.187)	0.599(0.113)	0.814(0.012)	0.574(0.214)
Uni-Mol	0.509(0.047)	0.614(0.023)	0.795(0.035)	0.561(0.054)
POM	0.557(0.029)	0.561(0.002)	0.771(0.018)	0.506(0.033)
NOSE	**0.636(0.047)**	**0.534(0.033)**	**0.846(0.012)**	**0.416(0.054)**

Table 34: Mixture prediction tasks: Pleasantness.

Method	Thresholds (Abraham)			
	R^2 ↑	MAE ↓	Pearson ↑	MSE ↓
Adapter (Desc: 10.00M, Rec: 10.00M)	0.161(0.616)	1.074(0.421)	0.768(0.049)	1.875(1.375)
Adapter (Desc: 29.17M, Rec: 10.00M)	0.331(0.179)	0.986(0.148)	0.742(0.042)	1.494(0.401)
Inter	0.338(0.446)	0.978(0.374)	0.782(0.063)	1.478(0.995)
Hard + Soft λ=1.0	0.440(0.072)	0.884(0.085)	0.749(0.006)	1.252(0.162)
Receptor only	0.444(0.173)	0.909(0.173)	0.791(0.041)	1.242(0.387)
Hard + Soft λ=0.5	0.480(0.037)	0.887(0.034)	0.771(0.020)	1.162(0.083)
Only Hard	0.449(0.261)	0.891(0.246)	0.810(0.030)	1.232(0.582)
Hard + Soft λ=0.1	0.538(0.097)	0.827(0.088)	0.764(0.056)	1.032(0.216)
Only Soft λ=2.0	0.515(0.050)	0.868(0.031)	0.795(0.008)	1.083(0.111)
No Orthogonal	0.506(0.254)	0.849(0.226)	0.804(0.022)	1.104(0.567)
Molecule only	0.581(0.064)	0.770(0.091)	0.779(0.047)	0.936(0.142)
Inter + Weak	0.570(0.100)	0.792(0.131)	0.813(0.004)	0.961(0.223)
Inter + Intra	0.582(0.140)	0.779(0.158)	0.802(0.031)	0.934(0.313)
Description only	0.608(0.026)	0.746(0.038)	0.816(0.025)	0.875(0.058)
NOSE	**0.652(0.072)**	**0.711(0.083)**	**0.836(0.026)**	**0.778(0.161)**

Table 35: Ablation study: Basic perceptual attribute prediction: Thresholds.

Method	Pleasantness (Keller)			
	R^2 ↑	MAE ↓	Pearson ↑	MSE ↓
No Orthogonal	0.343(0.078)	7.818(0.487)	0.622(0.046)	114.722(13.697)
Hard + Soft λ=1.0	0.354(0.055)	8.079(0.355)	0.623(0.040)	112.703(9.565)
Molecule only	-0.071(0.204)	10.576(1.354)	0.680(0.017)	186.939(35.551)
Description only	0.366(0.110)	7.959(0.877)	0.647(0.063)	110.652(19.235)
Receptor only	0.368(0.104)	8.025(0.690)	0.636(0.102)	110.308(18.106)
Inter + Weak	0.383(0.090)	7.751(0.422)	0.654(0.063)	107.715(15.682)
Inter + Intra	0.401(0.066)	7.632(0.518)	0.648(0.039)	104.572(11.437)
Adapter (Desc: 10.00M, Rec: 10.00M)	0.378(0.117)	7.393(0.688)	0.669(0.030)	108.561(20.457)
Hard + Soft λ=0.5	0.398(0.018)	7.681(0.457)	0.663(0.025)	105.027(3.067)
Adapter (Desc: 29.17M, Rec: 10.00M)	0.420(0.039)	7.486(0.332)	0.652(0.032)	101.285(6.863)
Hard + Soft λ=0.1	0.432(0.075)	7.197(0.638)	0.662(0.062)	99.119(13.024)
Only Soft λ=2.0	0.441(0.104)	7.087(0.816)	0.671(0.084)	97.555(18.119)
Only Hard	0.468(0.097)	7.004(0.368)	0.704(0.077)	92.803(16.947)
NOSE	0.488(0.074)	6.911(0.609)	0.715(0.050)	89.280(12.934)
Inter	**0.520(0.155)**	**6.878(0.967)**	**0.734(0.096)**	**83.762(27.123)**

Table 36: Ablation study: Pleasantness prediction (Keller).

Method	Pleasantness (Sagar)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
Inter + Intra	-0.282(0.336)	0.511(0.047)	0.025(0.245)	0.384(0.101)
Inter	-0.219(0.214)	0.505(0.023)	-0.034(0.285)	0.365(0.064)
Adapter (Desc: 10.00M, Rec: 10.00M)	-0.208(0.152)	0.492(0.036)	-0.127(0.181)	0.362(0.046)
Only Hard	-0.169(0.027)	0.508(0.016)	0.085(0.158)	0.350(0.008)
No Orthogonal	-0.148(0.110)	0.490(0.022)	0.083(0.078)	0.344(0.033)
Hard + Soft λ=0.5	-0.110(0.163)	0.484(0.029)	0.011(0.166)	0.333(0.049)
Adapter (Desc: 29.17M, Rec: 10.00M)	-0.083(0.167)	0.477(0.031)	0.129(0.329)	0.325(0.050)
Hard + Soft λ=1.0	-0.108(0.374)	0.474(0.061)	0.298(0.156)	0.332(0.112)
Receptor only	-0.103(0.031)	0.467(0.021)	0.188(0.056)	0.331(0.009)
Hard + Soft λ=0.1	-0.104(0.379)	0.476(0.069)	0.362(0.099)	0.331(0.114)
Inter + Weak	0.001(0.100)	0.476(0.022)	0.229(0.154)	0.299(0.030)
Description only	-0.009(0.105)	0.463(0.011)	0.209(0.122)	0.302(0.032)
Molecule only	0.030(0.170)	0.456(0.039)	0.144(0.335)	0.291(0.051)
Only Soft λ=2.0	0.064(0.099)	0.451(0.021)	0.223(0.249)	0.281(0.030)
NOSE	**0.105(0.064)**	**0.447(0.011)**	**0.397(0.020)**	**0.268(0.019)**

Table 37: Ablation study: Basic perceptual attribute prediction: Pleasantness (Sagar).

Method	Intensity (Keller)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
Molecule only	-0.470(0.317)	17.189(1.927)	0.270(0.031)	470.285(101.497)
Inter	-0.224(0.096)	16.201(0.767)	0.206(0.042)	391.690(30.734)
Hard + Soft λ=0.5	-0.097(0.141)	15.088(0.689)	0.246(0.074)	351.066(44.974)
Adapter (Desc: 29.17M, Rec: 10.00M)	-0.103(0.194)	14.797(1.517)	0.304(0.090)	352.751(62.021)
Hard + Soft λ=1.0	-0.024(0.165)	14.839(1.500)	0.322(0.117)	327.603(52.872)
Hard + Soft λ=0.1	0.020(0.052)	14.348(0.501)	0.320(0.036)	313.497(16.783)
Only Hard	-0.015(0.068)	14.513(0.815)	0.336(0.025)	324.564(21.719)
Adapter (Desc: 10.00M, Rec: 10.00M)	0.023(0.085)	14.074(0.633)	0.315(0.055)	312.679(27.280)
Description only	0.026(0.013)	14.231(0.442)	0.321(0.029)	311.741(4.179)
Receptor only	0.033(0.133)	13.952(1.122)	0.323(0.119)	309.369(42.499)
Inter + Weak	0.043(0.091)	14.029(0.819)	0.329(0.075)	306.240(29.163)
No Orthogonal	0.050(0.100)	14.234(0.574)	0.355(0.077)	303.921(32.127)
Only Soft λ=2.0	0.052(0.066)	13.684(0.673)	0.347(0.054)	303.291(21.067)
Inter + Intra	0.060(0.032)	14.207(0.238)	0.381(0.042)	300.820(10.274)
NOSE	**0.156(0.020)**	**12.932(0.230)**	**0.418(0.010)**	**269.902(6.526)**

Table 38: Ablation study: Basic perceptual attribute prediction: Intensity (Keller).

Method	Intensity (Sagar)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $\uparrow$	MSE $\downarrow$
Inter	-0.389(0.244)	0.455(0.034)	0.360(0.132)	0.283(0.050)
Only Soft λ=2.0	-0.253(0.355)	0.455(0.054)	0.347(0.180)	0.256(0.072)
Molecule only	-0.317(0.102)	0.442(0.022)	0.372(0.051)	0.269(0.021)
Hard + Soft λ=0.1	-0.112(0.065)	0.404(0.040)	0.273(0.166)	0.227(0.013)
No Orthogonal	-0.138(0.253)	0.398(0.072)	0.308(0.108)	0.232(0.052)
Adapter (Desc: 29.17M, Rec: 10.00M)	-0.177(0.132)	0.423(0.020)	0.430(0.087)	0.240(0.027)
Hard + Soft λ=0.5	-0.104(0.222)	0.423(0.045)	0.429(0.113)	0.225(0.045)
Description only	-0.072(0.142)	0.390(0.059)	0.378(0.173)	0.219(0.029)
Receptor only	-0.027(0.045)	0.384(0.017)	0.361(0.152)	0.210(0.009)
Hard + Soft λ=1.0	-0.027(0.184)	0.375(0.053)	0.377(0.076)	0.209(0.038)
Only Hard	-0.021(0.162)	0.391(0.020)	0.417(0.068)	0.208(0.033)
Inter + Weak	0.081(0.003)	**0.347(0.021)**	0.327(0.039)	0.188(0.001)
Adapter (Desc: 10.00M, Rec: 10.00M)	0.020(0.143)	0.381(0.049)	0.435(0.089)	0.200(0.029)
NOSE	0.120(0.109)	0.355(0.041)	0.468(0.075)	0.179(0.022)
Inter + Intra	**0.177(0.202)**	0.350(0.050)	**0.541(0.073)**	**0.168(0.041)**

Table 39: Ablation study: Basic perceptual attribute prediction: Intensity (Sagar).

Method	Intensity (Ravia)			
	R^2 ↑	MAE ↓	Pearson ↑	MSE ↓
Molecule only	-0.407(0.352)	15.176(1.993)	0.307(0.024)	390.301(97.745)
Adapter (Desc: 29.17M, Rec: 10.00M)	0.069(0.123)	11.873(0.576)	0.202(0.288)	258.179(34.039)
Receptor only	0.058(0.135)	11.988(0.719)	0.347(0.134)	261.213(37.444)
Inter + Intra	0.044(0.142)	12.107(0.601)	0.462(0.007)	265.066(39.293)
Adapter (Desc: 10.00M, Rec: 10.00M)	0.079(0.074)	12.012(0.130)	0.426(0.119)	255.522(20.592)
Hard + Soft λ=0.1	0.082(0.131)	11.788(0.637)	0.421(0.093)	254.596(36.293)
Only Soft λ=2.0	0.113(0.121)	11.761(0.608)	0.312(0.238)	245.860(33.608)
Hard + Soft λ=0.5	0.090(0.112)	11.746(0.577)	0.477(0.008)	252.269(30.948)
Inter	0.137(0.047)	11.712(0.133)	0.428(0.024)	239.377(13.032)
Hard + Soft λ=1.0	0.118(0.136)	11.667(0.772)	0.441(0.044)	244.606(37.752)
No Orthogonal	0.155(0.107)	11.550(0.548)	0.468(0.036)	234.259(29.573)
Only Hard	0.168(0.046)	11.564(0.407)	0.476(0.047)	230.751(12.671)
Description only	0.207(0.026)	**11.056(0.206)**	0.476(0.018)	219.985(7.194)
Inter + Weak	0.210(0.067)	11.205(0.815)	0.480(0.055)	219.046(18.540)
NOSE	**0.220(0.025)**	11.078(0.351)	**0.485(0.031)**	**216.326(6.824)**

Table 40: Ablation study: Basic perceptual attribute prediction: Intensity (Ravia).

Method	GS-LF Multi-label Multi-class Classification		
	AUC ↑	AUPRC ↑	MCC ↑
No Orthogonal	0.874(0.001)	0.344(0.007)	0.220(0.025)
Adapter (Desc: 29.17M, Rec: 10.00M)	0.875(0.003)	0.344(0.005)	0.250(0.024)
Only Hard	0.874(0.002)	0.342(0.002)	0.257(0.003)
Only Soft λ=2.0	0.874(0.001)	0.347(0.002)	0.237(0.016)
Inter + Weak	0.875(0.001)	0.345(0.004)	0.241(0.027)
Receptor only	0.875(0.002)	0.346(0.008)	0.244(0.019)
Molecule only	0.873(0.001)	0.347(0.005)	0.262(0.020)
Description only	0.876(0.001)	0.347(0.004)	0.241(0.010)
Hard + Soft λ=1.0	0.875(0.002)	0.345(0.003)	0.250(0.014)
Hard + Soft λ=0.5	0.874(0.004)	0.351(0.003)	0.256(0.018)
Hard + Soft λ=0.1	0.874(0.000)	0.348(0.003)	0.263(0.005)
Adapter (Desc: 10.00M, Rec: 10.00M)	0.874(0.004)	0.348(0.009)	0.261(0.010)
Inter + Intra	0.874(0.004)	0.351(0.003)	0.264(0.013)
NOSE	0.876(0.001)	0.351(0.002)	**0.268(0.010)**
Inter	**0.877(0.001)**	**0.354(0.004)**	0.260(0.014)

Table 41: Ablation study: semantic description prediction on GS-LF dataset.

Method	Multi-label Regression (Keller)			
	R^2 ↑	MAE ↓	Pearson r ↑	MSE ↓
Receptor only	-0.093(0.060)	6.369(0.226)	0.233(0.061)	79.899(3.491)
Description only	-0.060(0.062)	6.271(0.129)	0.217(0.074)	77.697(3.140)
Hard + Soft λ=0.5	-0.042(0.059)	6.089(0.350)	0.167(0.074)	75.739(6.335)
Molecule only	-0.178(0.060)	6.741(0.209)	0.330(0.050)	90.373(5.967)
Hard + Soft λ=1.0	-0.034(0.038)	6.136(0.141)	0.243(0.097)	75.666(4.221)
Inter	-0.010(0.043)	6.022(0.116)	0.191(0.053)	73.195(2.996)
Only Hard	-0.038(0.069)	5.993(0.084)	0.246(0.047)	75.320(4.707)
Inter + Weak	0.000(0.016)	5.882(0.183)	0.243(0.071)	71.656(0.863)
No Orthogonal	0.001(0.016)	6.069(0.244)	0.265(0.106)	71.798(3.617)
Only Soft λ=2.0	0.006(0.023)	5.977(0.147)	0.248(0.064)	71.690(1.685)
Adapter (Desc: 29.17M, Rec: 10.00M)	0.026(0.018)	5.866(0.176)	0.253(0.029)	69.822(1.679)
Inter + Intra	0.016(0.033)	5.760(0.083)	0.258(0.011)	70.365(2.269)
Hard + Soft λ=0.1	0.028(0.031)	5.815(0.269)	0.254(0.052)	69.549(3.590)
Adapter (Desc: 10.00M, Rec: 10.00M)	0.028(0.020)	**5.757(0.259)**	0.277(0.081)	69.049(2.518)
NOSE	**0.075(0.040)**	5.862(0.225)	**0.348(0.060)**	**67.161(4.161)**

Table 42: Ablation study: semantic description prediction on Keller dataset.

Method	Multi-label Regression (Sagar)			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $r \uparrow$	MSE $\downarrow$
Inter	-0.554(0.082)	0.356(0.020)	0.023(0.152)	0.239(0.015)
Inter + Weak	-0.389(0.012)	0.343(0.001)	0.007(0.108)	0.233(0.006)
Molecule only	-0.677(0.218)	0.355(0.009)	0.116(0.042)	0.252(0.002)
Hard + Soft λ=1.0	-0.334(0.161)	0.336(0.013)	-0.050(0.069)	0.225(0.014)
Inter + Intra	-0.364(0.158)	0.339(0.017)	0.042(0.176)	0.226(0.015)
Hard + Soft λ=0.5	-0.388(0.176)	0.339(0.014)	-0.024(0.085)	0.224(0.015)
Hard + Soft λ=0.1	-0.423(0.228)	0.332(0.022)	0.007(0.148)	0.224(0.018)
No Orthogonal	-0.329(0.131)	0.334(0.011)	-0.082(0.038)	0.224(0.012)
NOSE	-0.305(0.106)	0.343(0.017)	0.123(0.068)	0.225(0.017)
Receptor only	-0.338(0.152)	0.329(0.011)	0.075(0.056)	0.227(0.009)
Description only	-0.381(0.267)	0.330(0.022)	0.120(0.079)	0.222(0.030)
Adapter (Desc: 10.00M, Rec: 10.00M)	-0.298(0.113)	0.329(0.006)	0.075(0.152)	0.224(0.007)
Only Hard	-0.276(0.062)	0.334(0.012)	**0.169(0.179)**	0.224(0.016)
Adapter (Desc: 29.17M, Rec: 10.00M)	-0.277(0.071)	0.329(0.010)	0.022(0.216)	**0.211(0.018)**
Only Soft λ=2.0	**-0.241(0.080)**	**0.327(0.005)**	0.142(0.143)	0.212(0.002)

Table 43: Ablation study: semantic description prediction on Sagar dataset.

Method	Mixture Intensity			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $r \uparrow$	MSE $\downarrow$
Receptor only	0.224(0.085)	0.379(0.010)	0.582(0.012)	0.215(0.023)
Hard + Soft λ=0.1	0.258(0.095)	0.360(0.028)	0.530(0.093)	0.205(0.026)
Adapter (Desc: 10.00M, Rec: 10.00M)	0.263(0.063)	0.369(0.023)	0.575(0.029)	0.204(0.018)
Inter + Weak	0.285(0.049)	0.362(0.006)	0.566(0.061)	0.198(0.013)
Inter + Intra	0.285(0.117)	0.365(0.037)	0.576(0.050)	0.198(0.032)
Description only	0.285(0.086)	0.361(0.017)	0.549(0.070)	0.198(0.024)
Only Hard	0.291(0.081)	0.359(0.032)	0.586(0.042)	0.196(0.023)
Adapter (Desc: 29.17M, Rec: 10.00M)	0.335(0.098)	0.356(0.029)	0.611(0.041)	0.184(0.027)
Molecule only	0.319(0.106)	0.355(0.028)	0.637(0.069)	0.188(0.029)
Hard + Soft λ=0.5	0.356(0.046)	0.342(0.017)	0.624(0.039)	0.178(0.013)
Only Soft λ=2.0	0.360(0.072)	0.331(0.019)	0.603(0.062)	0.177(0.020)
No Orthogonal	0.360(0.069)	0.347(0.014)	0.617(0.056)	0.177(0.019)
Hard + Soft λ=1.0	0.387(0.094)	0.331(0.021)	0.632(0.069)	0.169(0.026)
Inter	0.385(0.067)	**0.328(0.017)**	0.638(0.055)	0.170(0.019)
NOSE	**0.389(0.052)**	0.333(0.021)	**0.657(0.029)**	**0.169(0.014)**

Table 44: Ablation study: mixture prediction on intensity.

Method	Mixture Pleasantness			
	$R^2 \uparrow$	MAE $\downarrow$	Pearson $r \uparrow$	MSE $\downarrow$
Adapter (Desc: 10.00M, Rec: 10.00M)	0.404(0.260)	0.671(0.133)	0.781(0.068)	0.681(0.298)
Hard + Soft λ=0.5	0.509(0.053)	0.610(0.042)	0.753(0.021)	0.561(0.060)
Molecule only	0.509(0.047)	0.614(0.023)	0.795(0.035)	0.561(0.054)
Adapter (Desc: 29.17M, Rec: 10.00M)	0.521(0.127)	0.591(0.066)	0.790(0.032)	0.547(0.145)
Hard + Soft λ=0.1	0.529(0.151)	0.582(0.091)	0.786(0.028)	0.538(0.172)
Only Soft λ=2.0	0.549(0.018)	0.603(0.013)	0.792(0.021)	0.516(0.021)
Inter	0.558(0.108)	0.586(0.061)	0.788(0.063)	0.505(0.123)
Only Hard	0.547(0.083)	0.582(0.029)	0.806(0.040)	0.518(0.094)
Description only	0.555(0.050)	0.574(0.028)	0.798(0.020)	0.508(0.058)
Hard + Soft λ=1.0	0.574(0.087)	0.572(0.044)	0.777(0.054)	0.487(0.099)
Inter + Weak	0.566(0.106)	0.565(0.075)	0.788(0.062)	0.496(0.121)
No Orthogonal	0.562(0.088)	0.577(0.067)	0.807(0.029)	0.501(0.101)
Inter + Intra	0.606(0.036)	0.557(0.021)	0.821(0.030)	0.450(0.041)
Receptor only	**0.648(0.047)**	0.536(0.026)	0.836(0.039)	**0.402(0.054)**
NOSE	0.636(0.047)	**0.534(0.033)**	**0.846(0.012)**	0.416(0.054)

Table 45: Ablation study: mixture prediction on pleasantness.