Optimizing Learning Across Multimodal Transfer Features for Modeling Olfactory Perception

Daniel Shin*
Stanford University and Sony AI
USA
dshin7@stanford.edu

Gao Pei*
Nara Institute of Science and Technology
Japan
gao.pei@naist.ac.jp

Priyadarshini Kumari†
Sony AI
USA
priyadarshini.kumari@sony.com

Tarek R. Besold
Sony AI
Spain
tarek.besold@sony.com

ABSTRACT

For humans and other animals, the sense of smell provides crucial information in many situations of everyday life. Still, the study of olfactory perception has received only limited attention outside of the biological sciences. From an AI perspective, the complexity of the interactions between olfactory receptors and volatile molecules and the scarcity of comprehensive olfactory datasets, present unique challenges in this sensory domain. Previous works have explored the relationship between molecular structure and odor descriptors using fully supervised training approaches. However, these methods are data-intensive and poorly generalize due to labeled data scarcity, particularly for rare-class samples.

Our study partially tackles the challenges of data scarcity and label skewness through multimodal transfer learning. We investigate the potential of large molecular foundation models trained on extensive unlabeled molecular data to effectively model olfactory perception. Additionally, we explore the integration of different molecular representations, including molecular graphs and text-based SMILES encodings, to achieve data efficiency and generalization of the learned model, particularly on sparsely represented classes. By leveraging complementary representations, we aim to learn robust perceptual features of odorants. However, we observe that traditional methods of combining modalities do not yield substantial gains in high-dimensional skewed label spaces. To address this challenge, we introduce a novel *label-balancer* technique specifically designed for high-dimensional multi-label and multi-modal training. The label-balancer technique distributes learning objectives across modalities to optimize collaboratively for distinct subsets of labels. Our results suggest that multi-modal transfer features learned using the label-balancer technique are more effective and robust, surpassing the capabilities of traditional uni- or multi-modal approaches, particularly on rare-class samples.

CCS CONCEPTS

• **Computing methodologies** → **Transfer learning**; **Semi-supervised learning settings**.

KEYWORDS

Multimodal transfer learning, foundation models, perception modelling, olfactory perception

ACM Reference Format:
Daniel Shin, Gao Pei, Priyadarshini Kumari, and Tarek R. Besold. 2018. Optimizing Learning Across Multimodal Transfer Features for Modeling Olfactory Perception. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The human sense of smell plays a crucial role in many domains, including food and flavor perception, perfumery, assistive technology, healthcare, and increasingly also in multimodal user interface design. Despite its significance, olfactory perception has received relatively limited scientific attention outside of the biological sciences. This is largely due to several domain-specific challenges unique to this sensory domain, such as the complex interactions between hundreds of olfactory receptors and volatile molecules and the scarcity of comprehensive olfactory datasets.

While modeling olfactory perception is still in its early stages, machine learning has emerged as a promising approach for addressing various complex problems in a neighboring field, namely chemistry: drug discovery [16] and protein folding [23] are only two of several examples. The enormous success of transformer and foundation models in the vision and the NLP domains, such as BERT [8], GPT [35], DALL-E [37], and T5 [36], has inspired the development of large molecular models like SMILES transformer [19], MG-BERT [52], and ChemBERT [6] to solve complex biomedical [23] and biochemical [12] problems. Using unsupervised or self-supervised methods, these models learn molecular fingerprints by pretraining sequence-to-sequence language models on SMILES data [22, 27]. SMILES ('*simplified molecular-input line-entry system*, [48]) is a text-based standard representation for molecules

*Both authors contributed equally to this research.
†Corresponding author.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

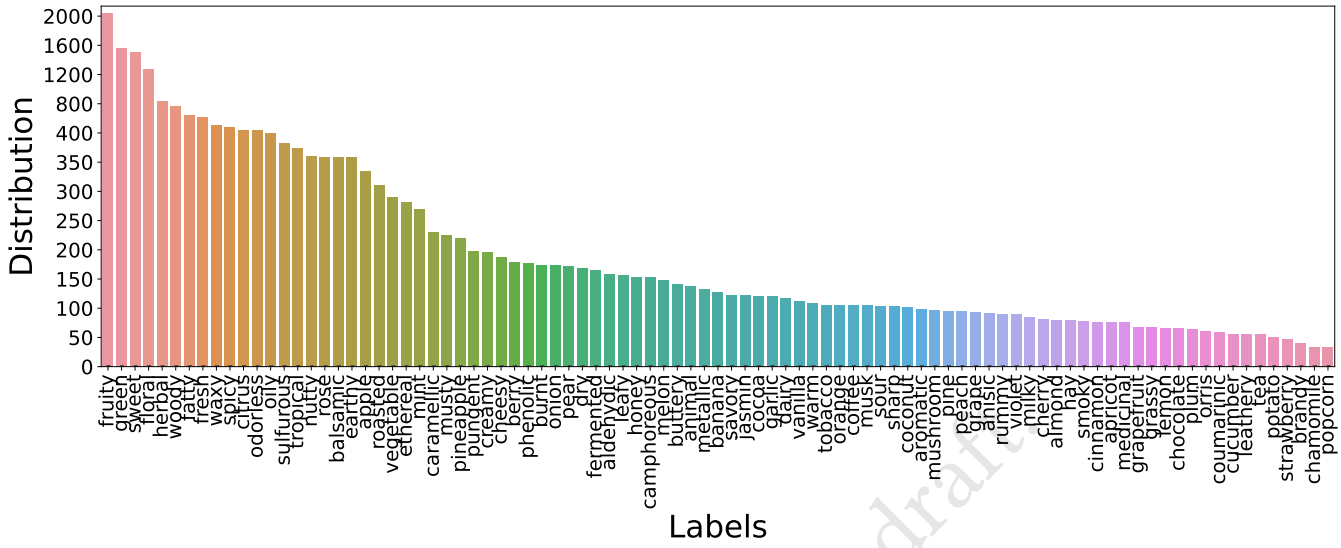


Figure 1: Distribution of perceptual descriptors of odorants on the entire dataset. Few descriptors, such as "fruity", and "sweet", are more often used to describe odor perception than other descriptors.

that's commonly used in computational chemistry. While the effectiveness of molecular foundation models such as ChemBERT and SMILES transformer has been extensively investigated in the domains of drug discovery and quantitative structure-property relationship (QSPR) prediction, their potential application for smell perception, also known as quantitative structure-odor relationship (QSOR) prediction, remains unexplored in prior research.

Recently, the QSOR problem has been approached using fully supervised training [14, 24, 25, 40, 42, 43]. Keller *et al.* [25] conducted an empirical study to investigate the physical properties of molecules that evoke specific smells such as "floral" or "pungent". Roberto *et al.* [42] proposed a distinct set of physico-chemical features that contribute to different smell perceptions, highlighting the role of sulfur atoms in evoking pungent smells. The most recent work, by Benjamin and Brian *et al.* [29, 40] employed a graph neural network (GNN) trained on molecular graphs to model odor perception. While olfactory perception has been studied long before, most classical approaches [18, 38, 39] rely on empirical studies to establish the relationship between molecular structure and odor descriptors. Despite these efforts, the precise connection between molecular structure and olfactory perception remains unclear. Furthermore, all of these fully-supervised methods are data-intensive, posing challenges in acquiring sufficient training data. The largest publicly available olfactory perceptual dataset, Goodscent [11], contains only 4626 labeled samples. Even a model trained on the entire dataset from scratch through a fully-supervised method still performs poorly, particularly on sparsely represented classes. Similar to other olfactory datasets, the label distribution of the Goodscent dataset is highly skewed, as illustrated in Figure 1. Certain odor descriptors, such as "fruity" and "sweet", are used more frequently than others, such as "tea" and "strawberry".

In this work, we address the problems of *data scarcity* and *label skewness* by leveraging multimodal transfer learning. Specifically, we investigate how large molecular foundation models trained on extensive *unlabeled* molecular data such as PubChem and ZINC [22, 27] can be potentially used to model olfactory perception effectively. Furthermore, we explore how combining different modalities–including (a) molecular graphs that capture the symmetry and orientation of atomic systems and (b) SMILES, a sequential text encoding of chemical formulae that enables the utilization of language models–contributes effectively to developing a data-efficient perceptual model. Unlike conventional multimodal learning approaches, a naive fusion of different modalities turns ineffective when dealing with high-dimensional and highly class-imbalanced data. To address this challenge, we introduce *label-balancer*, a technique for high-dimensional multi-label and multi-modal training frameworks. Our label-balancer technique distributes learning objectives across different modalities, allowing different models to optimize collaboratively for distinct subsets of labels. Our approach leads to improved generalization and overall performance compared to single-modality training or traditional multi-modality fusion approaches [3]. The performance gain from label-balancer is especially pronounced on rare-class samples.

Our main contributions are

(1) We introduce a data-efficient perceptual model through the utilization of multimodal transfer learning. We show that transfer features derived from pre-trained molecular foundation models are highly effective in perception modeling, even without prior training on perceptual labels. Remarkably, our method achieves comparable performance using only 20% of the available labeled data. This results in a substantial reduction of data requirements by 75% compared to non-transfer learning approaches.
(2) Additionally, we explore how different modality molecular representations contribute to olfactory perception modeling.

To address the problem of skewed label distribution, we introduce the label-balancer technique that improves model generalization and performance without any additional computational cost or training data requirements.

Finally, we conduct a comprehensive evaluation of our method, demonstrating its performance in comparison to prior approaches across diverse experimental scenarios. Benefiting from pre-training on large unlabeled data and combining two modalities, our framework trained via MolCLR [47] and SMILES-transformer [19] demonstrates better performance on olfactory perception modeling tasks in comparison to prior supervised learning methods. Our code can be found at https://github.com/priyadarshini-sony/multimodal-olfaction.

2 RELATED WORK

The relevant work can be broadly classified into three categories: olfactory perceptual model, transfer learning in olfaction, multimodal perception learning. We review representative techniques in each category and discuss the unique aspects of our approach compared to existing methods.

2.1 Olfactory perceptual model

Several studies utilize machine learning empowered tools to model olfactory perception [17, 33, 43]. Olfactory perception is based on perceived chemical stimuli, which are associated with complex physicochemical parameters of chemicals. Earlier studies investigated the relationships between the odor characteristics of chemicals and their physicochemical parameters using linear modeling approaches, including principal component analysis (PCA) and non-negative matrix factorization (NMF) [4, 26]. However, considering the fundamentally nonlinear nature of the biological olfactory system, we should question the suitability of these linear modeling techniques for accurately modeling olfactory perception. Nozaki *et al.* [33] utilize nonlinear dimensionality reduction on mass spectra data as inputs and use the language modeling method word2vec to predict odor characters of chemicals.

In addition to the information from mass spectrometry, subsequent studies incorporated additional chemical structure information as explanatory variable to improve the accuracy. Traditional hand-crafted molecular representations such as Dragon [42] and Mordred [31] are characterized by fixed-length vectors representing different physical and chemical properties of molecules. Gutierrez *et al.* [17] predicted up to 70 olfactory perceptual descriptors using chemoinformatic features generated by Dragon. Without using cheminformatics features, Tran *et al.* [43] hypothesized that chemicals play the role of ligands with 3D spatial structures to olfactory receptors and, therefore, can be learned using convolutional neural networks. They trained a convolutional auto-encoder, called DeepNose, to learn the mapping between a low-dimensional 3D spatial representation of molecules and human perceptual responses. Most recently, Sanchez-Lengeling *et al.* [40] trained a graph neural network to predict the relationship between a molecule's structure and its smell. The graph embeddings capture meaningful structures on both a local and global scale, which is useful in downstream QSOR tasks. Lee *et al.* [29] extends Sanchez-Lengeling *et al.* [40]'s work by employing a GNN to generate a Principal Odor Map (POM) that preserves and represents known perceptual relationships and enables odor quality prediction for novel odorants. However, none of these prior works explore multimodal transfer learning to address the problem of data efficiency and performance generalization, which are inherent in the olfactory domain.

2.2 Transfer learning in olfaction

Unlike the olfactory perception task (QSOR prediction), several fields in the biomedical and biochemical domains explore the utility of large molecular foundation models for a wide range of tasks. Using transfer learning, chemical language models have demonstrated their capability to learn specific chemical features from a much smaller training set. Several works develop large language models similar to BERT through self-supervised training on SMILES sequences (SMILES-BERT [46], MOLBERT [10], Bio-Bert [30], ChemBERTa [7], and ChemBERTa-2 [1]). After pretraining, these models are fine-tuned for their respective downstream tasks. On the other hand, the seq2seq model is also proposed to provide effective vector representations by leveraging a large pool of unlabeled data. SMILES2Vec is an interpretable general-purpose deep neural network for predicting various chemical properties, such as toxicity, activity, solubility, and solvation energy [15]. Among the large pretrained seq2seq models, Seq3seq [51]is the first semi-supervised learning model for molecular property prediction. It utilizes an Encoder-Decoder structure which can provide a strong molecular representation using a huge training data pool containing a mixture of both unlabeled and labeled molecules.

SMILES Transformer using a Transformer-based seq2seq architecture is another noteworthy approach introduced by Honda *et al.* [20]. It works well for the defined downstream predictive task, especially demonstrating improved performance in small data settings. The question of whether large language models, such as GPT-3, trained on non-chemical corpora, can acquire meaningful knowledge in the field of chemistry has also been investigated in a recent study [49].

Besides chemical language models, molecular graphs have been widely used for pretraining strategies. Hu *et al.* [21] propose to train a GNN model with context prediction or attribute masking self-supervised tasks. In the context prediction approach, a binary classifier is employed to determine whether a specific atom environment corresponds to a particular context graph. On the other hand, attribute masking involves masking random nodes, and the objective is to predict their attributes, such as atom type. Wang *et al.* [47] extend the pre-training with an alternative strategy based on contrastive learning. Benefiting from the different augmentation strategies they used, the fine-tuned MOLCLR (Molecular Contrastive Learning of Representations via Graph Neural Networks) model achieved state-of-the-art performance on various chemical tasks, including molecular property prediction. Despite several efforts in diverse domains, none of the prior works explore the utility of pre-trained transfer features for QSOR tasks.

2.3 Multi-modal perception learning

As human beings, our perception of the environment is shaped by the information we gather through multimodal multisensory

clues. A learning agent that aims to replicate human-like capabilities should also possess the ability to comprehend and generate information across different modalities. In order to learn representations of multimodal data, Silva*et al.* [41] propose to learn common encoded features using multimodal VAE (MVAE). Another approach, the Multimodal Factorization Model (MFM) [44], proposes the factorization of the multimodal representation into separate, independent representations. Vasco *et al.* [45] proposed a hierarchical design, called MUSE, to learn a hierarchical multimodal representation, beginning with low-level modality-specific representations from raw observation data and ending with a high-level multimodal representation encoding joint-modality information.

In the perception domain, vision and audio constitute the major part of multi-modal perception learning [3]. Chen *et al.* [5] propose a Vision-Audio-Language Omni-peRception pretraining model (VALOR) for multi-modal understanding and generation. Experiments show that VALOR can learn strong multimodal correlations and be generalized to various downstream tasks. In addition to visual and auditory, individuals that interact with the physical world, such as robots, will benefit from a fine-grained tactile perception of objects and surfaces. Gao *et al.* [13] propose a method of classifying surfaces with haptic adjectives from both visual and physical interaction data such as friction and vibration signals. Kumari *et al.* [34] proposed a deep neural network-based model of tactile perception that projects multiple sets of signals into the perceptual embedding space such as haptically-similar material surfaces are placed closer to each other. Richard *et al.* [50]'s work is similar to our study in showing the effectiveness of deep pre-trained features for visual perception modeling tasks. However, to the best of our knowledge, our work is the first to investigate the potential of incorporating multiple modalities and transfer features in the olfactory perception domain. This unique approach has the potential to provide valuable insights into the complex interplay between chemical features and human olfactory perception, opening up new avenues for understanding and improving odor perception models.

3 METHOD

In this section, we describe the process of extracting deep features from the molecular foundation model and calibrating them for the smell perception task. Next, we discuss the method of combining various modalities and training our multimodal framework in a highly dimensional and highly skewed label space, utilizing the label balancer technique. We begin by describing some standard molecular representations that are commonly used in machine learning applications.

The most commonly used molecular features for perceptual tasks are Dragon [42] and Mordred [31]. They are a collection of several types of molecular information in tabular form, describing physical or chemical properties of a molecule, such as the atom density, the number of carbon or sulfur atoms, and the acid/base count. Mordred features, being open-sourced, are more widely used in prior studies. Other representations include molecular graphs, which capture atomic system symmetry and orientation, and SMILES, a sequential text encoding of chemical formulae that enables the use of language models.

3.1 Perceptually-calibrated Transfer Features

SMILES-based perceptual features: In order to generate pre-trained deep features, we leverage SMILES-transfomer [32] which has been trained on $83M$ molecules from the PubChem [27] repository (i.e., one of the largest repositories of molecules, consisting of comprehensive information on molecular structure and properties). However, none of the prior approaches make use of this vast unlabeled dataset and large language models for the QSOR task that involves learning a mapping function between the molecule's structure and its smell perception. The SMILES transformer (shown in Figure 2) has a standard transformer architecture with six encoder and decoder layers and eight attention heads and is trained on a self-supervised task involving SMILES-IUPAC translation. IUPAC is an alternative text-based representation of molecular structure, describing similar aspects of molecular structure as SMILES but using a different nomenclature. During the training process, batches of 96 molecular string pairs are used, and the Adam optimization algorithm is applied with an initial learning rate of $1e^{-3}$. The learning rate follows a cosine function within each epoch, decreasing by two orders of magnitude after completing half a period. The training is performed over 83M molecules, lasting for three epochs.

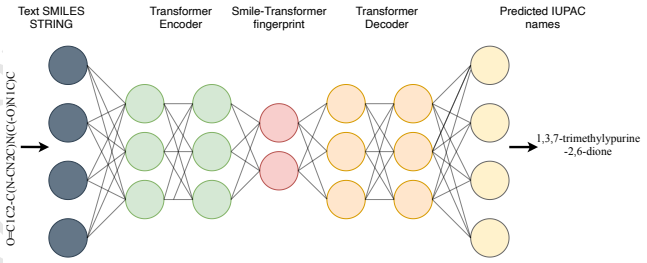


Figure 2: A depiction of SMILES-transformer trained through the self-supervised SMILES-IUPAC translation task. The learned embedded features are obtained from the encoder layer shown in red.

The intermediate features obtained from the pre-trained network effectively capture the information shared across both SMILES and IUPAC representations. To further refine these transfer features for olfactory perception, we perform perceptual calibration by fine-tuning the MLP head using supervision derived from olfactory perceptual descriptors. This learning technique leads to learning a more refined and optimized perceptual space, with higher weights for perceptually-relevant features and lower weights for less relevant ones. It is important to note that the SMILES-transfomer [32] is not trained using perceptual labels. Our approach only fine-tunes the MLP head to facilitate the downstream task of predicting QSORs. Our experimental results demonstrate that transfer features, even without explicit optimization for the perceptual task, are remarkably effective for the QSOR task. The transfer features require significantly less labeled data than the existing supervised approaches to achieve comparable performance.

Graph-based perceptual features: Although the SMILES-based representation effectively encodes sequential and structural properties, it fails to capture the crucial molecular topology. Given the

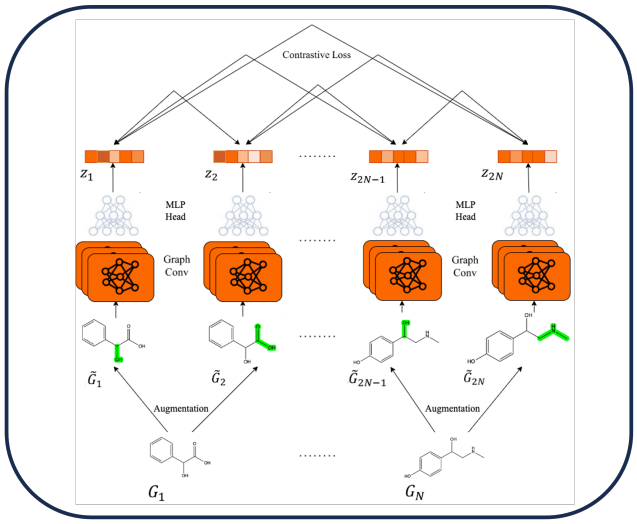


Figure 3: The MolCLR framework optimized for olfactory perception task. The green mask shows the removed subgraph resulting from the graph augmentation technique. The final layer embedding features, denoted as z_i, are fine-tuned for the downstream QSOR task.

vastness of the chemical space, it becomes challenging for any single molecular representation to generalize across a wide range of molecules. To have better coverage of the representational space, we explore an alternative modality representation, a molecular graph, which adequately captures the topology and structural orientation of molecules. Similar to SMILES, we employ a pre-trained model, MolCLR [47], trained on 10M PubChem molecules through self-supervised contrastive loss. MolCLR defines the self-supervised task using three different graph augmentation techniques: atom masking, bond deletion, and subgraph removal. The positive pairs constitute a molecule and its corresponding augmented molecule graph, while any two different molecules form negative pairs. Similar to the SMILES-transformer, our graph framework consists of pre-trained MolCLR and MLP head, and we fine-tune the MLP head for downstream QSOR tasks using perceptual labels. MolCLR uses a 5-layer graph convolution [28] with ReLU activation as the GNN backbone, incorporating modifications from Hu *et al.* [21] to support edge features. Graph-level readout is performed through average pooling, producing a 512-dimensional molecular representation. The NT-Xent loss is optimized using the Adam optimizer with weight decay 10e-5. The model is trained for a total of 50 epochs with a batch size of 512.

3.2 Multimodal Representation

Next, we explore how combining the graph with text-based molecular representations helps learn more effective perceptual features. Our method draws inspiration from ensemble learning approaches. Combining diverse modalities or models usually improves the performance of machine learning methods. This improvement becomes more pronounced when the features or models are dissimilar from each other, as they contribute uniquely to the learning process. Multimodal fusion offers several advantages – a) multimodal information may offer complementary information for the defined learning task, b) multimodal learning can be viewed as an ensemble learning approach where multiple models optimize for the same downstream task, resulting in improved and robust performance, and c) certain modalities may be more expensive to obtain than others, in which case multimodal learning can still operate even in the absence of one or a few modalities. We begin by investigating the effectiveness of classical fusion methods by optimizing uni-modal SMILE transfer features z_S, and graph transfer features z_G, individually as well as jointly using static fusion approaches such as concatenation $v_S \parallel z_G$, element-wise sum $z_S \oplus v_G$, and product $z_S \odot z_G$. The final embedding features, after combining the modalities using an element-wise sum, can be expressed as follows:

$$z_M = f_M(f_S(z_S) \oplus f_G(z_G)) \tag{1}$$

$$\tag{2}$$

Here, f_S and f_G represent the MLP heads for SMILE-transformer and MolCLR, respectively. f_M refers to the final linear layer that combines different modality features with optimal weights based on their perceptual relevance.

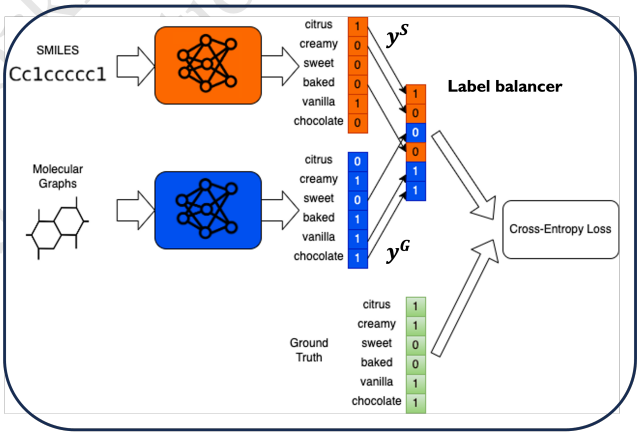


Figure 4: The label balancer minimizes the impact of the skewed dataset and helps achieve better performance and generalization on all class samples. The sets y^S and y^G denote subsets of labels optimized by SMILES-transformer and MolCLR, respectively. After each training round of weight updates, both modality features are combined using a static fusion approach.

Label-balancer: We observe that the final multimodal features are less effective due to the high correlation between modalities. This correlation leads to a reduced amount of complementary information available to aid the models. Furthermore, training becomes even more challenging with a high-dimensional skewed label distribution. To address this problem, we introduce the label-balancer training technique, which mitigates overfitting and offers better generalization on rare-class test samples. The core idea is to distribute learning objectives across different modalities, enabling different models to optimize collaboratively for distinct subsets of

labels. In each training iteration, we use a random label division strategy where half of the labels are optimized using the SMILES transformer, and the remaining half are optimized using the MolCLR. Specifically, our objective function is

$$L_{ce} = -\sum_{i=1}^{L} logp_i^y \mathbb{1}_{y^S}(i) - logp_i^y \mathbb{1}_{y^G}(i) \quad (3)$$

Here y^S and y^G are complementary sets, denoting label subsets optimized by SMILES-transformer and MolCLR, respectively. The division of labels among different models enables each model to learn more effective features for the assigned labels. Moreover, by integrating diverse features learned from distinct models trained on different label sets, our method demonstrates improved generalization capability, which is particularly difficult to achieve with high-dimensional multi-label data. Our proposed training framework improves performance compared to uni-modal training and classical multi-modality fusion approaches [3]. We anticipate further improvement in the performance as we incorporate additional modalities into our training framework.

4 EVALUATION

We evaluate our framework through several experiments addressing three questions: (Q1) How effective are pre-trained molecular foundation models for modeling olfactory perception? (Q2) Does combining different molecular representations, such as molecular graphs and text-based SMILES, result in better perceptual features? (Q3) How effective is our multimodal training technique, label-balancer, compared to classical fusion techniques for high-dimensional and highly skewed multi-label spaces? We start by introducing the dataset, implementation details, and evaluation setup before discussing each question in subsequent subsections.

4.1 Dataset and Implementation

Dataset. In the olfactory domain, there are very few perception datasets. The commonly used datasets include the Dravenieks database [9], which comprises only 138 molecules described by a $131D$ dimensional perceptual label vector. Additionally, the Keller dataset [25] consists of 480 molecules, with non-expert provided $20D$ dimensional descriptors. Other notable datasets are the Goodscents dataset [11], containing 4626 molecules described by $668D$ dimensional descriptors, and the Leffingwell dataset [2], consisting of 3522 molecules described by $113D$ dimensional descriptors. The descriptor labels for both datasets, Goodscents and Leffingwell, are gathered from domain-experts and hence are less noisy. While the Dravenieks database [9] is too small to be effectively used in learning-based techniques, the Keller dataset suffers from noise and sparsity issues due to the labels being collected from non-experts. To generate a large-scale and clean dataset, after filtering out noisy labels and inconsistent molecules, we compiled a collection of 5595 molecules from the Goodscents and Leffingwell datasets, described by $91D$ dimensional perceptual descriptors. Even after cleaning out noisy labels, the final curated dataset has a skewed label distribution, where certain descriptors such as "fruity" are frequently used, while descriptors like "dairy" and "tea" are sparsely used. Moreover, the label set is also fine-grained, consisting of broad and commonly used descriptors like fruity to more specific labels such as apple, pear, pineapple, etc.

Implementation Details. We use the SMILES-transformer, consisting of six encoded and decoded layers [19] and MolCLR [47], consisting of a 5-layer graph convolution with ReLU activation. Both modality representations are fine-tuned using an MLP head with 512 neurons. We train the model using the cross-entropy loss, with Adam optimizer on a batch size of 32 for 5000 epochs. We evaluate the performance of our learned model using the AUROC (Area Under the Receiver Operating Characteristic) metric, which is commonly used for multilabel classification problems. We measure the model's performance by calculating the unweighted mean AUROC, which involves averaging the AUROC scores across all 91 odor descriptors and assigning equal weights to all descriptors to ensure unbiased performance comparison.

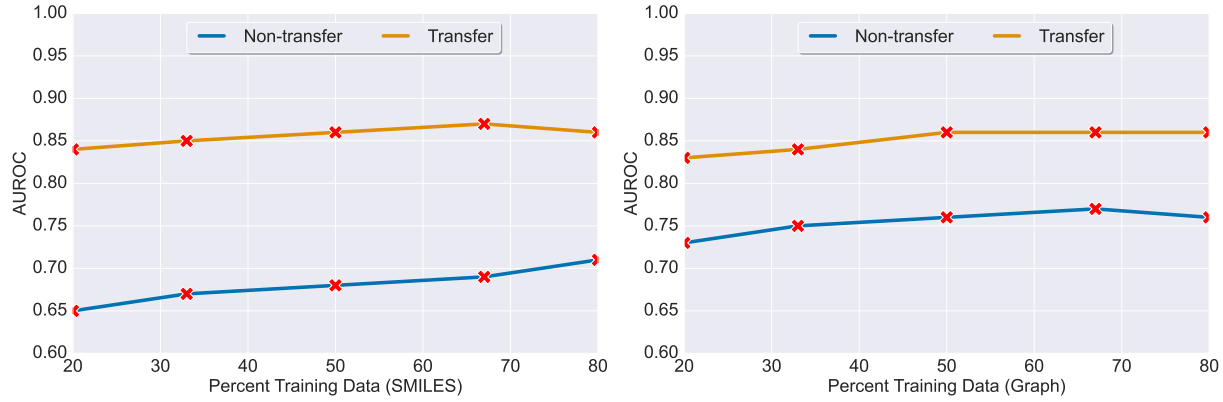
4.2 Results and Discussion

Q1: How effective are pre-trained molecular foundation models for modeling olfactory perception?

Figure 5a shows the performance comparison between perceptual features learned with and without using a pre-trained SMILES-transformer model. To obtain non-transfer features, we use state-of-the-art by zheng [53]. We train the SMILES-transformer [53] from scratch using our perceptual training data. To evaluate whether the model's performance is limited by the amount of data, we conduct training on progressively larger volumes of data from 20% to 80% of the whole dataset. The results demonstrate that the performance of the non-transfer model improves as more data is used, but it remains suboptimal even after utilizing 80% of the available training data. However, by leveraging the pre-trained model [19], our method significantly outperforms the state-of-the-art. Even with just 20% of the available labeled data, we obtain considerably higher AUROC compared to the performance achieved without transfer learning.

It is interesting to note that the pre-trained SMILES-transformer [19] is trained using a self-supervised task of SMILES-IUPAC translation, which does not have an obvious connection with smell perception. Remarkably, the transfer features learned from this self-supervised objective are perceptually effective and yield substantial performance gains with minimal perceptual supervision on only 20% of the data. Moreover, the computational overhead for learning weights for just one MLP head is significantly reduced compared to training the entire model from scratch. The rate of performance improvement with increasing training data is less pronounced in the case of transfer learning. This can be attributed to the diminished potential for improvement over the already rich and effective pre-trained features generated from millions of unlabeled samples.

Figure 5b depicts a similar comparison for the molecular graph. We learn non-transfer graph features using another state-of-the-art approach in the molecular graph domain by Sanchez-Lengeling *et al.* [40]. They utilize a graph neural network to model olfactory perception and evaluate their approach using a curated Goodscents dataset. As their data is not publicly available, we train their model on our data to demonstrate the performance of the learned non-transfer features. As shown in figure 5b, our method significantly outperforms



(a) (b)

Figure 5: Performance of perceptual features learned with and without transfer learning on (a) SMILES representation (left) (b) molecular graph representation (right) with an increasing amount of training data. For both representations, our approach, leveraging pre-trained features from molecular foundation models, outperforms state-of-the-art methods by a significant margin.

the state-of-the-art. However, it is worth noting that the performance of non-transfer features derived from the molecular graph is considerably superior to that of the SMILES representation. This observation aligns with what would can intuitively be expected, as the molecular graph is more effective in capturing key elements for modeling olfactory perception (such as the presence or absence of atoms, types of atomic bonds, orientation, and topology) than the text-based simpler representation provided by SMILES.

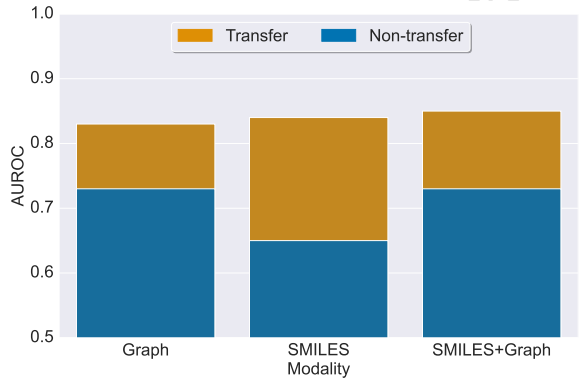


Figure 6: Ablation study to show the benefit of transfer learning on uni-modal and multi-modal representations. Transfer learning helps in both uni- and multi-modal features.

Next, we examine the benefits of transfer learning for both uni-modal and multi-modal features. Similar to the previous evaluations, we utilize state-of-the-art methods [40, 53] for the SMILES and graph representations, when learning non-transfer features. To derive multi-modal features, we combine the graph and SMILES representations using a traditional fusion method that involves element-wise summation. As shown in figure 6, we observe that the advantages of transfer learning extend beyond single-modality features. Even in the case of multi-modal features, significant performance gains are achieved through the use of transfer learning.

> **Takeaway 1** The transfer features acquired from molecular foundation models demonstrate remarkable **perceptual** effectiveness and robustness across diverse feature types and datasets.

Q2: Does combining different molecular representations, such as molecular graphs and text-based SMILES, result in better perceptual features?

Table 1: Performance of different uni- and multi-modal features for olfactory perception tasks. For the sensitivity analysis, we compare the performance of all the standard fusion variations, including element-wise sum $\oplus$, product $\odot$, and concatenation $\|$.

Features	Multimodal	Test AUROC
SMILES (S) [53]	✗	0.71
Graph (G) [40]	✗	0.76
MORDRED (M) [31]	✗	0.80
$S \oplus G$	✓	0.81
$S \oplus M$	✓	0.83
$G \oplus M$	✓	0.84
$S \oplus G \oplus M$	✓	0.81
$S \odot G \odot M$	✓	0.84
$S \,\|\| \, G \,\|\| \, M$	✓	0.84

Table 1 provides an overview of the performance of different uni- and multimodal features for olfactory perception tasks. For the single modality features, we compare our method with the relevant state-of-the-art approaches. In all cases, we train the model on 80% of the available data and test it on the remaining 20% of the data. The

results clearly demonstrate that while there is some improvement achieved by combining modalities, the gains are relatively less significant compared to the benefits observed from transfer features. Among all the standard fusion variations, such as element-wise sum, product, and concatenation, we observe the best performance with the concatenation and element-wise product features.

Upon further examination of the poor performance of multimodal features, we find that different modalities tend to make similar errors. Specifically, we observe that all modalities make accurate predictions on samples belonging to well-represented classes, while simultaneously making errors on samples from rare classes. Figure 7 shows the percentage of test samples where both SMILES and the graph model make correct or incorrect predictions. As depicted in the figure 7, there are only few samples on which both models make different predictions. This result suggests that the individual modality is not diverse enough to contribute complementary information to the overall performance of the combined model.

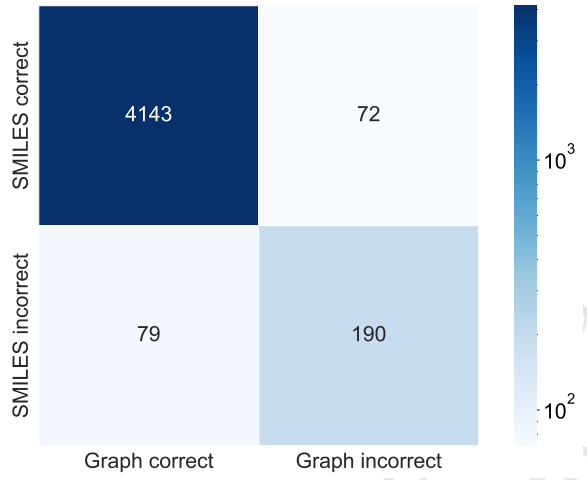


Figure 7: Distribution of test samples where both SMILES and graph models make (dis)similar predictions. There are only few samples where both models make different predictions.

> **Takeaway 2** The integration of graph-based features and SMILES representations results in limited improvements when compared to transfer learning methods due to a lack of complementary information.

Q3: How effective is our multimodal training technique, label-balancer, compared to classical fusion techniques for high-dimensional and highly skewed multi-label spaces?

Next, we evaluate the effectiveness of our proposed label balancer technique compared to classical fusion approaches for combining different modalities. To evaluate multimodality techniques, we train two models: one utilizes the classical element-wise sum and fine-tuning with an MLP head, while the other employs the label balancer technique to train the joint model, using both SMILES and graph inputs. We evaluate the performance of both models by comparing an average test AUROC value computed across all 91 perceptual descriptors. To assess the robustness of the label balancer technique, we conduct experiments on training data of varying sizes ranging from 5% to 80%. In the figure 8, the blue curve represents the performance of the combined model trained using the MLP head, while the yellow curve represents the performance using the label balancer technique. Notably, the label balancer technique consistently outperforms the classical fusion approach across all training subsets. This further emphasizes the effectiveness and robustness of our approach for multilabel multimodal training. While figure 8 demonstrates the effectiveness of the label balancer compared to the MLP head, the gain is normalized due to averaging AUROC across all label descriptors. Next, we explicitly demonstrate how the label balancer technique affects the performance of well and sparsely represented classes separately.

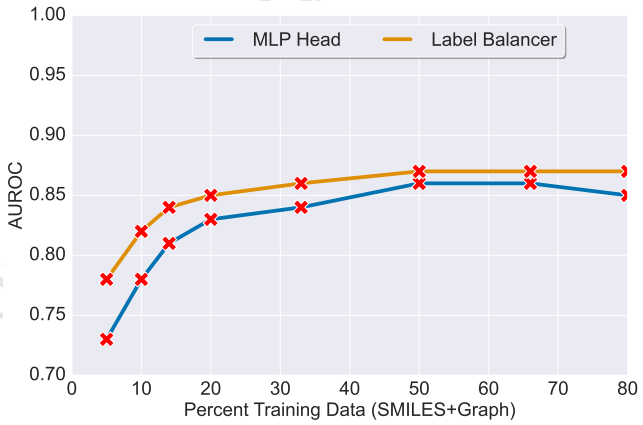


Figure 8: Performance comparison between classical multimodal fusion technique (MLP head) and label balancer. The result shows the average test AUROC value across all 91 perceptual descriptors. Our label balancer consistently outperforms the MLP head across all training dataset sizes.

We evaluate the performance of the MLP head and the label balancer on each class separately. For ease of visualization, we group all classes into clusters based on the sample density. In figure 9, the x-axis represents the clusters formed from all 91 descriptor classes, ranging from the most dense class (left-most) to the most sparse class (right-most).

For each cluster, we show the average performance gain achieved by the label balancer over the MLP head. As we see in the figure 9, the label balancer technique yields higher gains on sparsely represented classes compared to densely represented classes. The increasing trend from left to right validates our intuition that training with distributed objectives across different modalities helps in the generalization of the model for rare class samples. Furthermore, the performance on the dense class is already good, leaving less room for improvement for the label balancer technique. To see the performance gain for each class, please refer to table 3 in the appendix.

Finally, we examine the multimodal representation learned by our label balancer for combining SMILES and graph features. In order to visualize the learned embedding, we use t-SNE algorithms

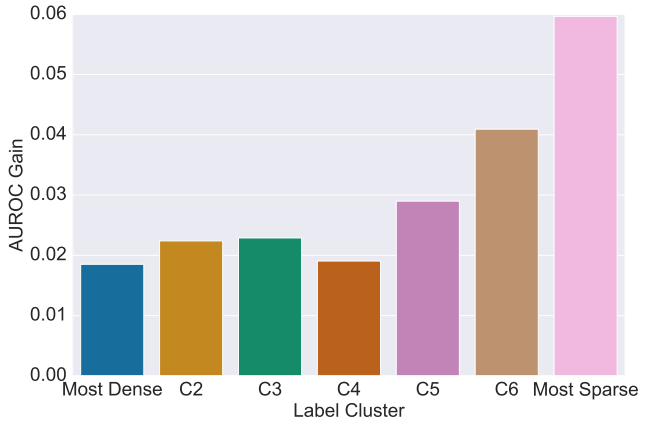


Figure 9: Performance gain by the label balancer over the MLP head approach on most-dense to most-sparse classes.

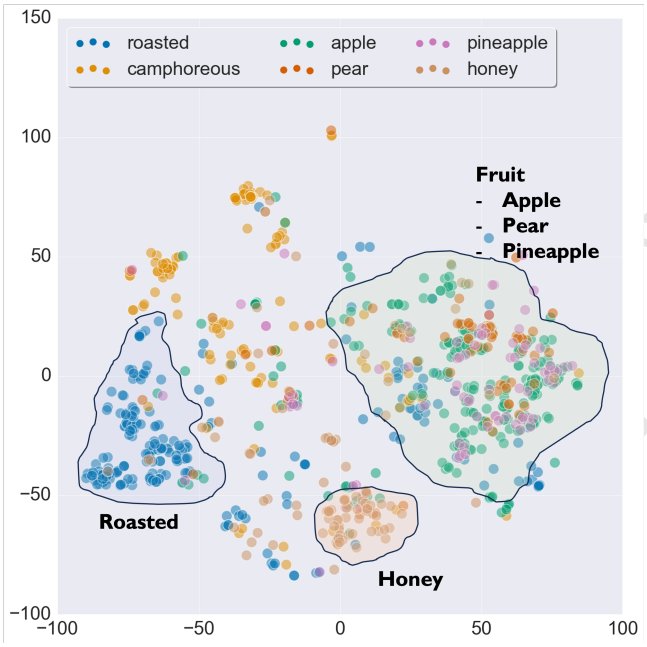


Figure 10: Visualization of molecular representation learned by our model via t-SNE. Representations are shown for the test samples on randomly selected labels, which are shown by different colors.

to project the learned representation of test samples onto a 2D space. We want to emphasize that each sample in our case has multiple labels, resulting in molecules belonging to different class clusters rather than a single one. As a result, we observe diffused clusters in the embedding space, in contrast to the tight clusters observed in multiclass problems where classes are mutually exclusive. Despite this, we still see perceptually similar classes appearing closer to each other than distinct ones. For instance, molecules that evoke fruit smells such as "apple", "pear", and "pineapple" form a cluster and are perceptual neighbors in the embedding space. Similarly, other flavors like "roasted" and "honey" are also grouped together. This observation suggests that our method captures the perceptual similarity between different flavors in a meaningful manner, despite the challenges posed by the multi-label nature of the problem.

Takeaway 3 The label balancer is effective in learning multimodal representation, particularly in high-dimensional and highly skewed multi-label spaces. It successfully learns perceptually meaningful representations and improves generalization, specifically for sparsely represented classes.

5 CONCLUSIONS

Our study addresses the challenges of data scarcity and label skewness in olfactory perception modeling by leveraging multimodal transfer learning. We demonstrate that the pre-trained molecular foundation models are effective in learning olfactory perception with minimal supervision. We address the data scarcity problem by leveraging pre-trained features and reducing the amount of data required for training by up to 75% compared to non-transfer learning approaches. We further investigate the effectiveness of different molecular representations and introduce the label-balancer technique to improve model generalization and performance in scenarios where the label space is high-dimensional and highly skewed. Experimental results on the largest publicly available olfactory perception dataset, Goodscents, validate that our method achieves both data efficiency and robust performance compared to state-of-the-art methods.

There are several interesting research directions to explore as future work. We would like to build a molecular foundation model, leveraging Mordred features similar to SMILES-transfer and MolCLR. We anticipate that incorporating additional modalities can further improve the performance of the label-balancer approach. We would like to explore the effectiveness of the label-balancer technique across a wider range of modalities and their combinations. Additionally, exploring novel approaches for multilabel and multimodal learning that exhibit robustness and generalization across different classes would be of great interest and crucial for understanding human smell perception.

REFERENCES

[1] Walid Ahmad, Elana Simon, Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. 2022. Chemberta-2: Towards chemical foundation models. *arXiv preprint arXiv:2209.01712* (2022).
[2] LEFFINGWELL ASSOCIATES. 2001. Database of Perfumery Materials & Performance. (2001). http://www.leffingwell.com/bacispmp.htm.
[3] Tadas Baltrušaitis, Chaitanya Ahuja, and Louis-Philippe Morency. 2018. Multimodal machine learning: A survey and taxonomy. *IEEE transactions on pattern analysis and machine intelligence* 41, 2 (2018), 423–443.
[4] Jason B Castro, Arvind Ramanathan, and Chakra S Chennubhotla. 2013. Categorical dimensions of human odor descriptor space revealed by non-negative matrix factorization. *PloS one* 8, 9 (2013), e73289.
[5] Sihan Chen, Xingjian He, Longteng Guo, Xinxin Zhu, Weining Wang, Jinhui Tang, and Jing Liu. 2023. Valor: Vision-audio-language omni-perception pretraining model and dataset. *arXiv preprint arXiv:2304.08345* (2023).
[6] Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. 2020. Chemberta: Large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885* (2020).
[7] Seyone Chithrananda, Gabriel Grand, and Bharath Ramsundar. 2020. Chemberta: Large-scale self-supervised pretraining for molecular property prediction. *arXiv preprint arXiv:2010.09885* (2020).

[8] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[9] Andrew Dravnieks et al. 1985. *Atlas of odor character profiles.*

[10] Benedek Fabian, Thomas Edlich, Héléna Gaspar, Marwin Segler, Joshua Meyers, Marco Fiscato, and Mohamed Ahmed. 2020. Molecular representation learning with language models and domain-relevant auxiliary tasks. *arXiv preprint arXiv:2011.13230* (2020).

[11] Fragrance Flavor. [n. d.]. Food, and Cosmetics Ingredients information. The Good Scents Company.

[12] Wenhao Gao and Connor W Coley. 2020. The synthesizability of molecules proposed by generative models. *Journal of chemical information and modeling* 60, 12 (2020), 5714–5723.

[13] Yang Gao, Lisa Anne Hendricks, Katherine J Kuchenbecker, and Trevor Darrell. 2016. Deep learning for tactile understanding from visual and haptic data. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 536–543.

[14] Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. 2017. Neural message passing for quantum chemistry. In *International conference on machine learning*. PMLR, 1263–1272.

[15] Garrett B Goh, Nathan O Hodas, Charles Siegel, and Abhinav Vishnu. 2017. Smiles2vec: An interpretable general-purpose deep neural network for predicting chemical properties. *arXiv preprint arXiv:1712.02034* (2017).

[16] Daria Grechishnikova. 2021. Transformer neural network for protein-specific de novo drug generation as a machine translation problem. *Scientific reports* 11, 1 (2021), 1–13.

[17] E Darío Gutiérrez, Amit Dhurandhar, Andreas Keller, Pablo Meyer, and Guillermo A Cecchi. 2018. Predicting natural language descriptions of monomolecular odorants. *Nature communications* 9, 1 (2018), 4979.

[18] Rafi Haddad, Rehan Khan, Yuji K Takahashi, Kensaku Mori, David Harel, and Noam Sobel. 2008. A metric for odorant comparison. *Nature methods* 5, 5 (2008), 425–429.

[19] Shion Honda, Shoi Shi, and Hiroki R Ueda. 2019. Smiles transformer: Pre-trained molecular fingerprint for low data drug discovery. *arXiv preprint arXiv:1911.04738* (2019).

[20] Shion Honda, Shoi Shi, and Hiroki R Ueda. 2019. Smiles transformer: Pre-trained molecular fingerprint for low data drug discovery. *arXiv preprint arXiv:1911.04738* (2019).

[21] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec. 2019. Strategies for pre-training graph neural networks. *arXiv preprint arXiv:1905.12265* (2019).

[22] John J Irwin and Brian K Shoichet. 2005. ZINC- a free database of commercially available compounds for virtual screening. *Journal of chemical information and modeling* 45, 1 (2005), 177–182.

[23] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. 2021. Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 7873 (2021), 583–589.

[24] Kathrin Kaeppler and Friedrich Mueller. 2013. Odor classification: a review of factors influencing perception-based odor arrangements. *Chemical senses* 38, 3 (2013), 189–209.

[25] Andreas Keller and Leslie B Vosshall. 2016. Olfactory perception of chemically diverse molecules. *BMC neuroscience* 17, 1 (2016), 1–17.

[26] Rehan M Khan, Chung-Hay Luk, Adeen Flinker, Amit Aggarwal, Hadas Lapid, Rafi Haddad, and Noam Sobel. 2007. Predicting odor pleasantness from odorant structure: pleasantness as a reflection of the physical world. *Journal of Neuroscience* 27, 37 (2007), 10015–10023.

[27] Sunghwan Kim, Paul A Thiessen, Evan E Bolton, Jie Chen, Gang Fu, Asta Gindulyte, Lianyi Han, Jane He, Siqian He, Benjamin A Shoemaker, et al. 2016. PubChem substance and compound databases. *Nucleic acids research* 44, D1 (2016), D1202–D1213.

[28] Thomas N Kipf and Max Welling. 2016. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).

[29] Brian K Lee, Emily J Mayhew, Benjamin Sanchez-Lengeling, Jennifer N Wei, Wesley W Qian, Kelsie Little, Matthew Andres, Britney B Nguyen, Theresa Moloy, Jane K Parker, et al. 2022. A Principal Odor Map Unifies Diverse Tasks in Human Olfactory Perception. *bioRxiv* (2022), 2022–09.

[30] Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics* 36, 4 (2020), 1234–1240.

[31] Hirotomo Moriwaki, Yu-Shi Tian, Norihito Kawashita, and Tatsuya Takagi. 2018. Mordred: a molecular descriptor calculator. *Journal of cheminformatics* 10, 1 (2018), 1–14.

[32] Paul Morris, Rachel St. Clair, William Edward Hahn, and Elan Barenholtz. 2020. Predicting binding from screening assays with transformer network embeddings. *Journal of Chemical Information and Modeling* 60, 9 (2020), 4191–4199.

[33] Yuji Nozaki and Takamichi Nakamoto. 2018. Predictive modeling for odor character of a chemical using machine learning combined with natural language processing. *PloS one* 13, 6 (2018), e0198475.

[34] Kumari Priyadarshini, Siddhartha Chaudhuri, and Subhasis Chaudhuri. 2019. PerceptNet: Learning Perceptual Similarity of Haptic Textures in Presence of Unorderable Triplets. In *IEEE World Haptics Conference (WHC)*.

[35] Alec Radford, Karthik Narasimhan, Tim Salimans, Ilya Sutskever, et al. 2018. Improving language understanding by generative pre-training. (2018).

[36] Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research* 21, 1 (2020), 5485–5551.

[37] Aditya Ramesh, Mikhail Pavlov, Gabriel Goh, Scott Gray, Chelsea Voss, Alec Radford, Mark Chen, and Ilya Sutskever. 2021. Zero-shot text-to-image generation. In *International Conference on Machine Learning*. PMLR, 8821–8831.

[38] 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 (2020), 118–123.

[39] Karen J Rossiter. 1996. Structure- odor relationships. *Chemical reviews* 96, 8 (1996), 3201–3240.

[40] 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* (2019).

[41] Rui Silva, Miguel Vasco, Francisco S Melo, Ana Paiva, and Manuela Veloso. 2019. Playing games in the dark: An approach for cross-modality transfer in reinforcement learning. *arXiv preprint arXiv:1911.12851* (2019).

[42] Roberto Todeschini and Viviana Consonni. 2009. *Molecular descriptors for chemoinformatics. 1. Alphabetical listing.* Wiley-VCH.

[43] 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*. PMLR, 6305–6314.

[44] Yao-Hung Hubert Tsai, Paul Pu Liang, Amir Zadeh, Louis-Philippe Morency, and Ruslan Salakhutdinov. 2018. Learning factorized multimodal representations. *arXiv preprint arXiv:1806.06176* (2018).

[45] Miguel Vasco, Hang Yin, Francisco S Melo, and Ana Paiva. 2021. How to sense the world: Leveraging hierarchy in multimodal perception for robust reinforcement learning agents. *arXiv preprint arXiv:2110.03608* (2021).

[46] Sheng Wang, Yuzhi Guo, Yuhong Wang, Hongmao Sun, and Junzhou Huang. 2019. SMILES-BERT: large scale unsupervised pre-training for molecular property prediction. In *Proceedings of the 10th ACM international conference on bioinformatics, computational biology and health informatics*. 429–436.

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

[48] 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 (1988), 31–36. https://doi.org/10.1021/ci00057a005

[49] Andrew D White, Glen M Hocky, Heta A Gandhi, Mehrad Ansari, Sam Cox, Geemi P Wellawatte, Subarna Sasmal, Ziyue Yang, Kangxin Liu, Yuvraj Singh, et al. 2022. Do large language models know chemistry? (2022).

[50] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 586–595.

[51] Xiaoyu Zhang, Sheng Wang, Feiyun Zhu, Zheng Xu, Yuhong Wang, and Junzhou Huang. 2018. Seq3seq fingerprint: towards end-to-end semi-supervised deep drug discovery. In *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics*. 404–413.

[52] Xiao-Chen Zhang, Cheng-Kun Wu, Zhi-Jiang Yang, Zhen-Xing Wu, Jia-Cai Yi, Chang-Yu Hsieh, Ting-Jun Hou, and Dong-Sheng Cao. 2021. MG-BERT: leveraging unsupervised atomic representation learning for molecular property prediction. *Briefings in bioinformatics* 22, 6 (2021), bbab152.

[53] Xiaofan Zheng, Yoichi Tomiura, and Kenshi Hayashi. 2022. Investigation of the structure-odor relationship using a Transformer model. *Journal of Cheminformatics* 14, 1 (2022), 88.

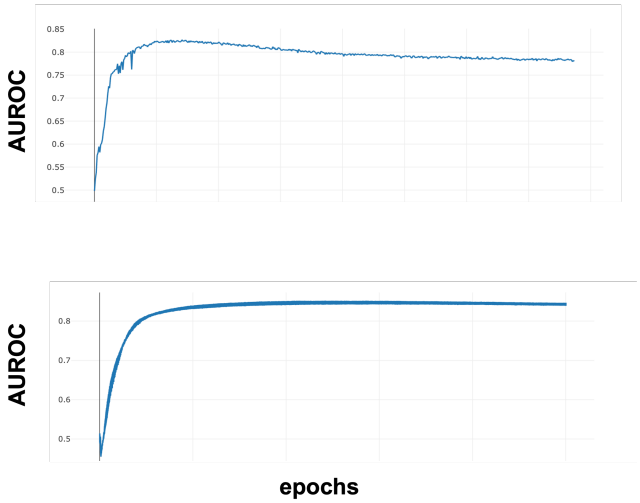


Figure 11: The upper figures depict the test AUROC of the MLP head fusion approach as the number of epochs increases, while the lower figure illustrates the same with the label balancer technique. In the first case, the performance initially improves but eventually starts to deteriorate, suggesting model overfitting. Conversely, the lower curve remains consistently robust throughout the training process.

A APPENDIX

A.1 Robustness of model learned using traditional MLP head fusion vs. label-balancer technique

A.2 Ablation study to show the contribution of multimodal and transfer learning for modeling olfactory perception

Table 2: Performance comparison with and without transfer or/and multimodal learning for modeling olfactory perception.

Features	Transfer	Multimodal	Test AUROC
SMILES (S) [53]	✗	✗	0.71
Graph (G) [40]	✗	✗	0.76
MORDRED (M) [31]	✗	✗	0.80
$S + G$	✗	✓	0.81
$S + M$	✗	✓	0.83
$G + M$	✗	✓	0.84
$S + G + M$	✗	✓	0.84
$S + G$ with Label Balancer	✓	✓	0.87

A.3 Performance comparisons of label balancer and MLP head on each class

Table 3: Performance comparison between MLP head and label balancer technique on each smell descriptor class.

Descriptors	Classical Fusion [⊕]	Label Balancer
aldehydic	0.89	0.94
almond	0.88	0.92
animal	0.68	0.85
anisic	0.81	0.78
apple	0.93	0.95
apricot	0.85	0.87
aromatic	0.69	0.74
balsamic	0.84	0.87
banana	0.89	0.89
berry	0.82	0.85
brandy	0.85	0.92
burnt	0.87	0.9
buttery	0.83	0.85
camphoreous	0.9	0.92
caramellic	0.85	0.88
chamomile	0.97	0.94
cheesy	0.82	0.87
cherry	0.87	0.88
chocolate	0.8	0.9
cinnamon	0.9	0.87
citrus	0.85	0.87
cocoa	0.87	0.92
coconut	0.76	0.86
coffee	0.93	0.94
coumarinic	0.88	0.93
creamy	0.74	0.84
cucumber	0.95	0.97
dairy	0.75	0.83
dry	0.76	0.68
earthy	0.71	0.78
ethereal	0.91	0.89
fatty	0.84	0.86
fermented	0.85	0.85
floral	0.87	0.87
fresh	0.73	0.76
fruity	0.84	0.86
garlic	0.99	0.98
grape	0.8	0.85
grapefruit	0.67	0.9
grassy	0.88	0.86
green	0.81	0.83
hay	0.78	0.77
herbal	0.72	0.79
honey	0.84	0.86
jasmin	0.87	0.87
leafy	0.76	0.77

Table 3: Performance comparison between MLP head and label balancer technique on each smell descriptor class.

Descriptors	Classical Fusion [⊕]	Label Balancer
leathery	0.73	0.78
lemon	0.8	0.79
medicinal	0.87	0.87
melon	0.91	0.93
metallic	0.71	0.72
milky	0.71	0.86
mint	0.86	0.89
mushroom	0.74	0.87
musk	0.96	0.98
musty	0.7	0.71
nutty	0.88	0.89
odorless	0.94	0.94
oily	0.76	0.81
onion	0.97	0.97
orange	0.84	0.94
orris	0.74	0.75
peach	0.76	0.88
pear	0.89	0.9
phenolic	0.88	0.89
pine	0.88	0.9
pineapple	0.87	0.88
plum	0.85	0.85
popcorn	0.99	0.99
potato	0.92	0.91
pungent	0.82	0.82
roasted	0.92	0.92
rose	0.87	0.92
rummy	0.81	0.83
savory	0.95	0.92
sharp	0.62	0.7
smoky	0.93	0.95
sour	0.76	0.79
spicy	0.76	0.79
strawberry	0.7	0.84
sulfurous	0.98	0.97
sweet	0.72	0.72
tea	0.61	0.71
tobacco	0.76	0.85
tropical	0.8	0.84
vanilla	0.91	0.92
vegetable	0.81	0.82
violet	0.86	0.89
warm	0.66	0.71
waxy	0.89	0.88
woody	0.84	0.83