COME: Dual Structure-Semantic Learning with Collaborative MoE for Universal Lesion Detection Across Heterogeneous Ultrasound Datasets

Lingyu Chen[1,*], Yawen Zeng[2,*], Yue Wang[1], Peng Wan[1], Guochen Ning[3], Hongen Liao[3,4], Daoqiang Zhang[1], Fang Chen[4,†]
[1]College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics
[2]ByteDance Inc. [3]Tsinghua University [4]Shanghai Jiaotong University
lingyucher@163.com, yawenzeng11@gmail.com, chenfang_bme@163.com

1. Overview

2. Limitations

We must also candidly acknowledge some limitations in our research, specifically: 1) Model: Since the MoE model is still under development, the current design is not yet cutting-edge. For instance, techniques like using LoRA for weight integration and training larger MoE models have not been fully implemented. 2) Application scenarios: Although our study has encompassed nearly all mainstream datasets, there is potential for further extension to other scenarios to verify generalizability. 3) Types of tasks: Our COME focuses on multi-source heterogeneous datasets. The next step is to expand its capabilities to comprehensively address a wider range of medical tasks.

3. More Details

3.1. Details of Dataset

To develop a universal model for heterogeneous ultrasound (US) datasets, we built a benchmark of 4 breast and 4 thyroid US datasets. These datasets come from different sources and exhibit significant domain differences, such as variations in shadow artifacts, speckle noise, grayscale levels, and anatomical structures, as shown in Fig. 2.

Due to strict collection conditions and reliance on expert doctors, many US datasets are imbalanced (see Fig. 1). However, the proposed COME architecture effectively overcomes this issue, as demonstrated in Table 2 of the main text.

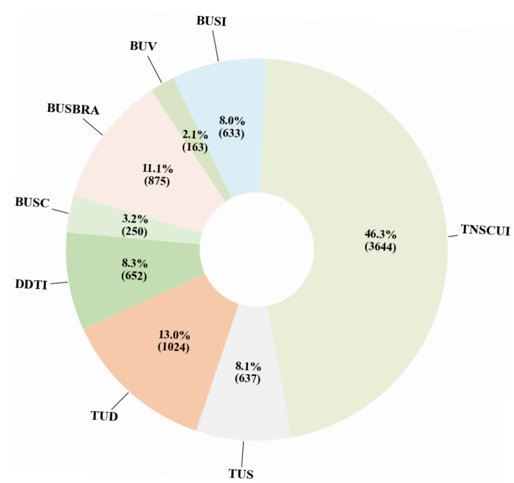

Figure 1. The dataset proportions and corresponding image counts in the benchmark.

3.2. Additional Qualitative Evaluation

In the main text, we select one sample per dataset for comparison. Here, Fig. 3 shows additional lesion detection examples, demonstrating that our structure-semantic learning-based COME model delivers robust performance on diverse US images and holds promise for real-world clinical applications.

3.3. Details of Ablations

In the main text, Figure 4 shows qualitative ablation visualizations. Here, Tab. 1 provides quantitative results that further demonstrate the effectiveness of each COME component.

3.4. Parameter Setting on Intra-organ Datasets

In the main text, we explore how the number of experts affects COME's performance on the inter-organ integrated dataset. Here, we evaluate its sensitivity on the intra-organ thyroid dataset (see Tab. 2). dataset-specific experts through

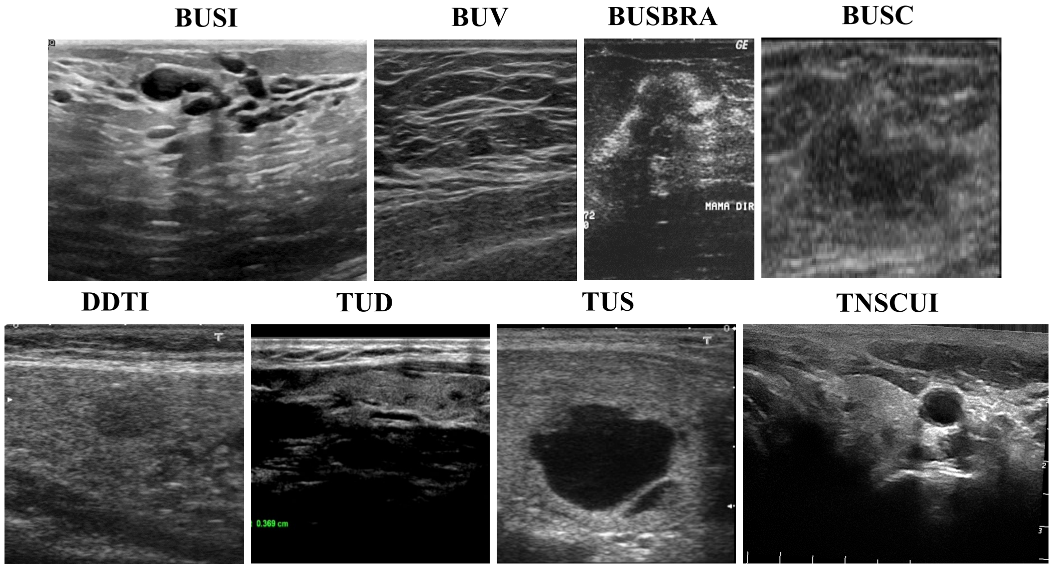

Figure 2. This paper constructs an integrated benchmark comprising eight heterogeneous breast and thyroid US datasets. And the distinct characteristics of each dataset pose challenges in building a universal analysis framework.

Table 1. Quantitative Performance of the ablation study.

Method	BUSI	BUV	BUSBRA	BUSC	DDTI	TUD	TUS	TNSCUI	Mean
STE	0.4628	0.6802	0.6517	0.7123	0.5008	0.6827	0.5307	0.6704	0.6115
SEE	0.3849	0.6570	0.5927	0.7006	0.5231	0.6859	0.5397	0.6795	0.5954
Dual Shared Experts(-DSE)	0.3853	0.6445	0.5509	0.6897	0.5173	0.6687	0.5189	0.6802	0.5819
Clustering	0.4587	0.7093	0.6590	0.7003	0.5335	0.6952	0.5772	0.6779	0.6264
Traceability Loss	0.4721	0.7211	0.6605	0.6913	0.5341	0.6960	**0.5826**	0.6981	0.6320
Our COME(Fine2Coarse)	**0.5159**	**0.8313**	**0.6719**	**0.7266**	**0.5371**	**0.7091**	0.5725	**0.7052**	**0.6587**

Table 2. Effect of the number of experts on the intra-organ (thyroid) integration dataset.

# Experts	TUD	TUS	DDTI	TNSCUI	Mean
2	0.5095	**0.7092**	**0.5794**	0.6921	0.6226
4	0.5594	0.6932	0.5750	**0.6935**	0.6303
8	0.5481	0.7032	0.5900	0.6875	**0.6322**
10	**0.5595**	0.6919	0.5580	0.6878	0.6243

traceability loss and heterogeneous architecture, effectively isolating source features and minimizing interference. To ensure comprehensive analysis, we specifically include a 2-expert configuration for the intra-organ dataset with four sources, demonstrating its viability under constrained conditions.

4. Feature Visualization during Clustering

The proposed COME achieved optimal performance with its Fine2Coarse clustering strategy. Fig. 4 presents additional samples illustrating the feature distributions.

Simultaneously, we demonstrate a multi-step clustering in Fig. 5.

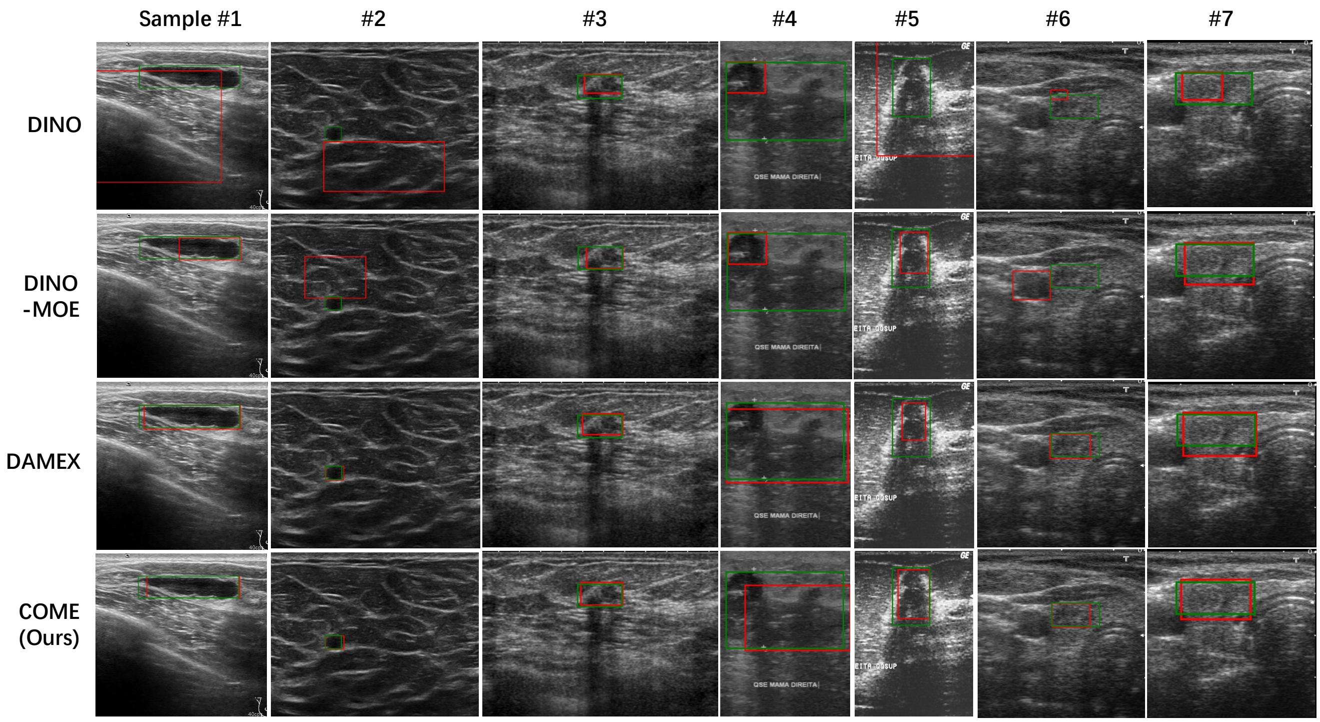

Figure 3. Additional lesion detection examples from the inter-organ integrated dataset.

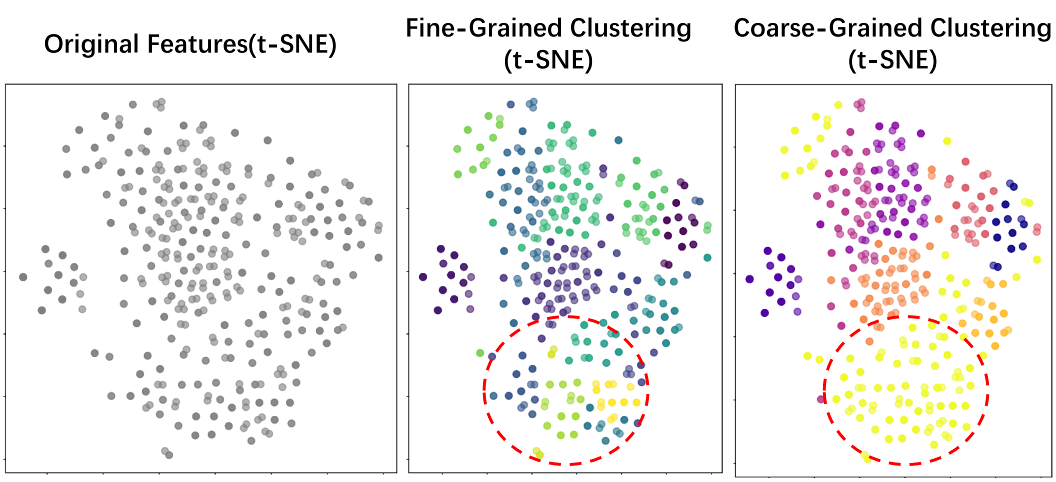

Figure 4. Feature visualization during training of the FineCoarse hierarchy clustering.

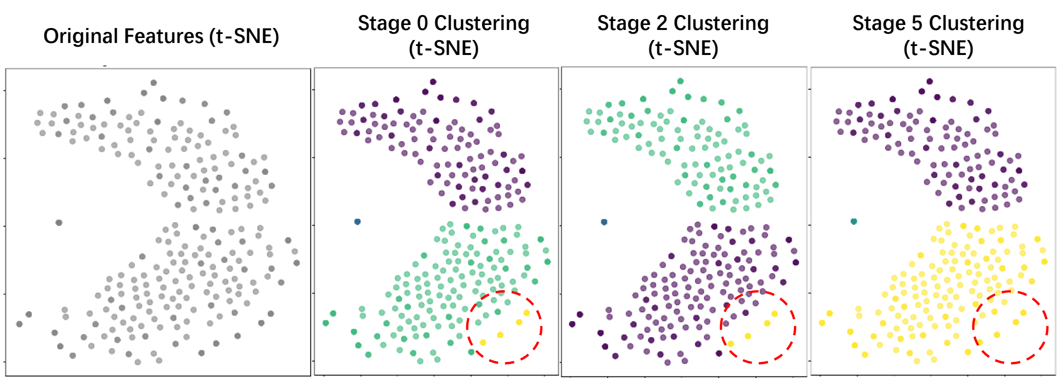

Figure 5. Feature visualization during training of COME's Multi-Step clustering.