TRACE: Training-time Report-guided and Clinically Ordered Concept Editing

Wentao Yue[1]†, Tianyou Lai[1]†, Jiayu Luo[1], Qingyu Mao[2], Ziying Wang[3], Zhenyuan Ning[3]*, Qilei Li[4]*

[1] Lanzhou University [2] Shenzhen University [3] Southern Medical University [4] Central China Normal University

† Equal contribution * Corresponding authors: jonnyning@foxmail.com; qilei.li@ccnu.edu.cn

Poster Fe-M12 · Nov. 13 – Main Conference · 16:00–18:00 · Versailles **ACM Multimedia 2026 (ACM MM '26) | Rio de Janeiro | Nov. 10–14, 2026**

Motivation and Main Idea

Clinical concepts support interpretable diagnosis

Shape, margin, orientation, echogenicity, posterior features, and calcification provide a clinically meaningful bridge from ultrasound appearance to benign or malignant decisions.

Reports provide valuable supervision but are unavailable at deployment

Structured reports contain expert concept descriptions during model development, yet they may be incomplete or entirely unavailable at test time. TRACE therefore uses reports only as privileged training-time supervision.

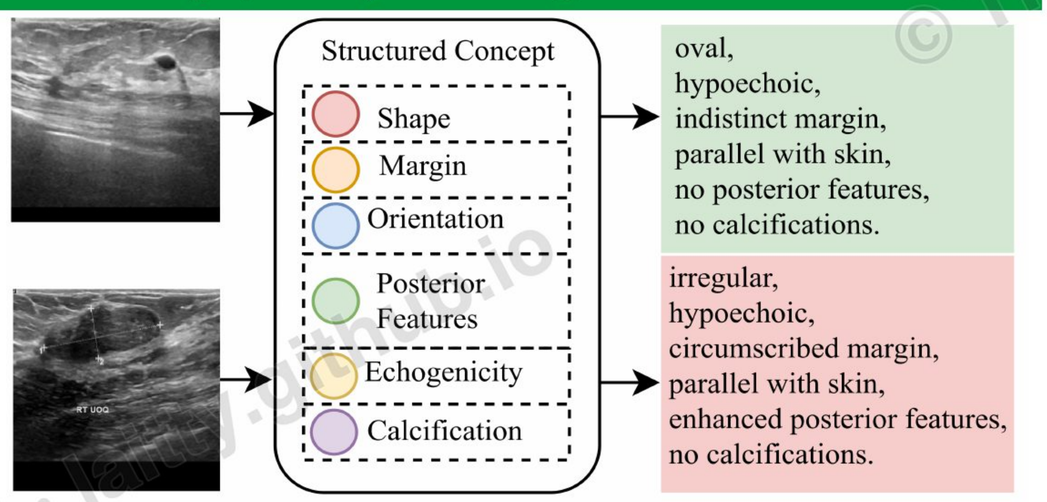
Coarse image-derived concepts require clinical correction

Noise, device variation, and domain shift can make initial concept predictions unstable. TRACE learns residual corrections in a malignancy-aware ordered space instead of predicting a new concept representation from scratch.

Main Idea

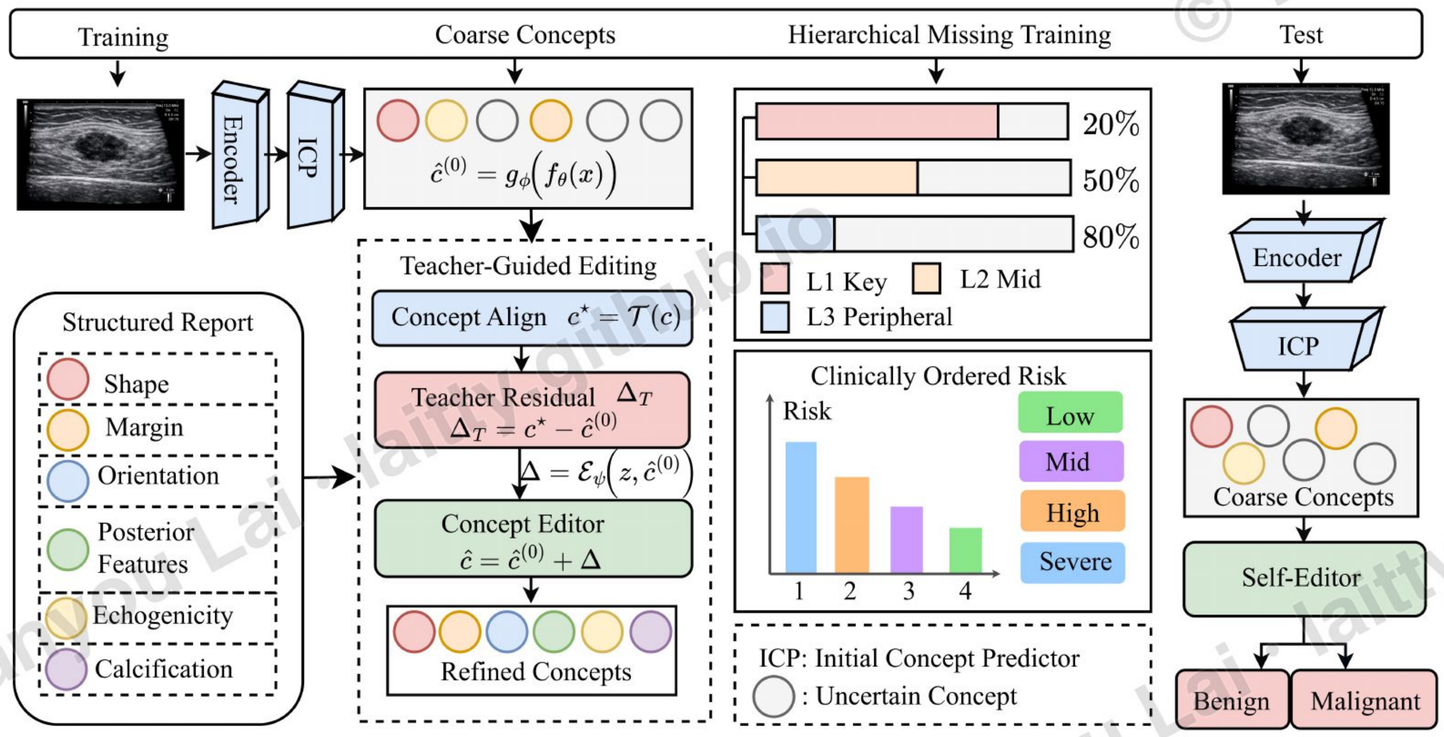
TRACE transfers report-derived clinical knowledge into an image-only self-editor: reports guide concept refinement during training, while diagnosis at inference requires only ultrasound images.

BUSC: Images, Concepts and Diagnostic Labels



BUSC: 1,872 BUSBRA + 647 BUSI lesion cases with six clinical concepts

TRACE Framework



Report-guided concept editing

The encoder and ICP first predict coarse concepts from the image (1). Report concepts are mapped to the same space, and their difference defines the teacher residual (2).

$$z = f\theta(x), \quad \hat{c}^{(0)} = g\varphi(z) \qquad (1)$$

$$c^T = T(c), \quad \Delta^T = c^T - \hat{c}^{(0)} \qquad (2)$$

Residual concept editing

The editor predicts a correction and adds it to the coarse concept state (3), preserving an interpretable edit path.

$$\Delta\psi = E\psi(z, \hat{c}^{(0)}), \quad \hat{c} = \hat{c}^{(0)} + \Delta\psi \qquad (3)$$

Strategic Concept Missing Training

Concepts are grouped by clinical importance. SCMT applies 20%, 50% and 80% missing rates and masks the teacher residual, forcing cross-concept reasoning (4).

$$r^L = 0.2 + 0.3(L-1), \quad L\in\{1,2,3\}$$
$$b^{(m)} \sim \text{Bernoulli}(1-r^{Lm}), \quad \Delta^T\text{mask} = b \odot \Delta^T \qquad (4)$$

Image-only inference

At test time the report branch is removed. The same editor uses image features and coarse concepts to self-edit before classification (5).

$$\Delta^s = E\psi(z, \hat{c}^{(0)}), \quad \hat{c} = \hat{c}^{(0)} + \Delta^s, \quad \hat{y} = h\omega(\hat{c}) \qquad (5)$$

Classification and concept-level supervision are optimized jointly in (6).

$$\mathcal{L} = \mathcal{L}cls + \lambda_1\mathcal{L}init + \lambda_2\mathcal{L}mask\text{-}edit + \lambda_3\mathcal{L}ord + \lambda_4\mathcal{L}cons \qquad (6)$$

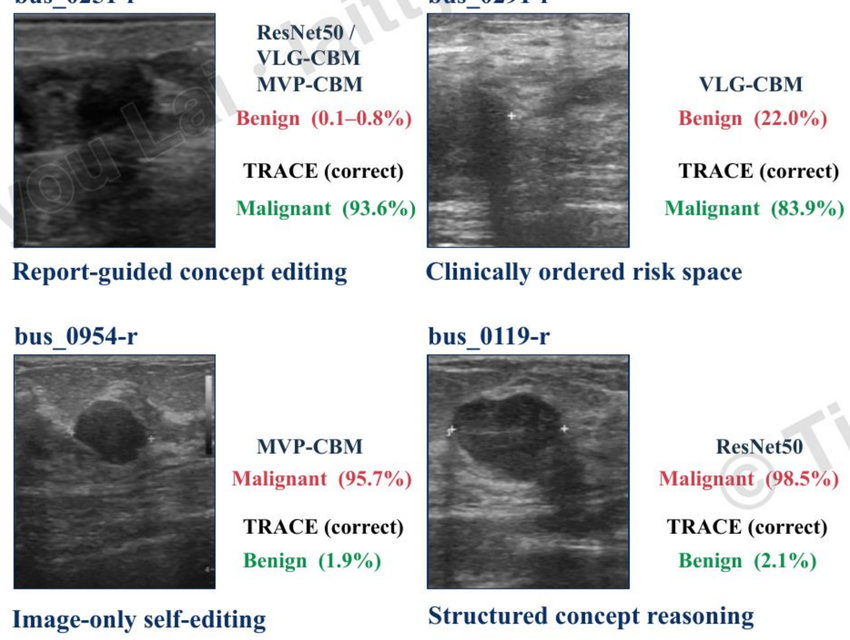
In-Domain Results

Five-fold CV · image-only inference · mean ± std (%) **Best in red; Second-best in blue**

Method	BUSC-BUSBRA AUC	BUSC-BUSBRA Accuracy	BUSC-BUSBRA F1	BUSC-BUSI647 AUC	BUSC-BUSI647 Accuracy	BUSC-BUSI647 F1
ResNet50	89.7 ± 1.5	82.6 ± 1.4	72.5 ± 2.7	94.5 ± 2.7	90.1 ± 2.6	83.8 ± 4.7
ViT-B/16	87.6 ± 1.8	82.1 ± 2.5	71.1 ± 4.2	94.4 ± 2.1	88.6 ± 3.5	80.9 ± 6.8
ProtoCaps	72.7 ± 3.7	71.8 ± 2.4	39.1 ± 7.5	80.3 ± 2.4	80.7 ± 1.4	65.8 ± 5.2
PCBM	88.6 ± 1.9	84.1 ± 1.8	73.8 ± 3.1	95.1 ± 2.2	90.1 ± 2.5	83.7 ± 4.7
MVP-CBM	88.9 ± 2.0	83.0 ± 2.1	74.9 ± 2.7	93.1 ± 2.7	87.8 ± 4.5	80.6 ± 6.9
VLG-CBM	88.3 ± 1.8	83.7 ± 2.4	73.3 ± 3.7	94.1 ± 2.8	91.0 ± 1.8	85.7 ± 2.9
Explicd	89.1 ± 2.5	83.3 ± 2.7	73.0 ± 5.0	93.9 ± 2.9	88.4 ± 1.6	82.2 ± 3.4
CLIP	55.3 ± 3.3	66.2 ± 1.1	11.1 ± 4.2	53.6 ± 2.0	67.4 ± 1.0	1.8 ± 3.6
TRACE	**91.6 ± 1.4**	**85.9 ± 1.3**	**77.6 ± 2.6**	**97.0 ± 3.0**	**93.8 ± 5.7**	**90.2 ± 8.3**

Visualization

Baselines fail; TRACE recovers the correct diagnosis

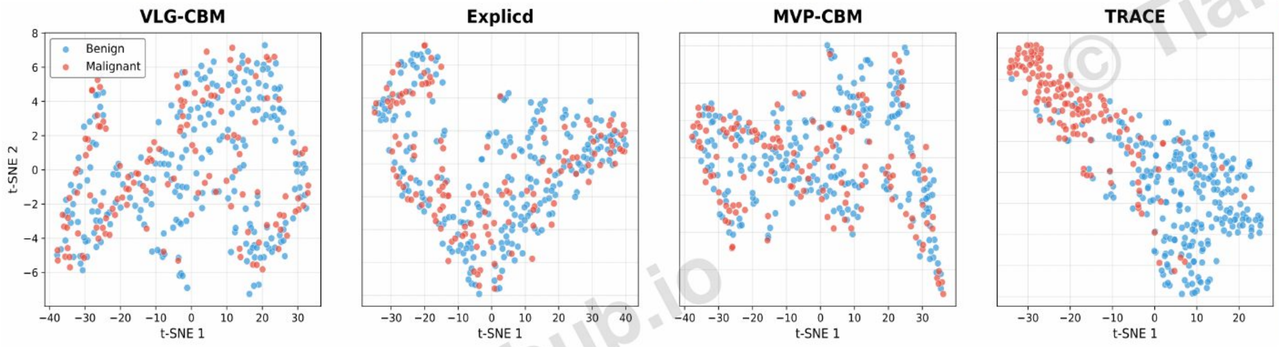


Zero-Shot Cross-Domain Transfer

TRACE, trained on BUSC-BUSBRA; no target fine-tuning.

Target domain	AUC (%)	Accuracy (%)
Ardakani	86.2 ± 1.8	83.6 ± 4.6
BUS_UC	66.1 ± 2.7	60.4 ± 2.3
BrEaST	82.2 ± 1.9	75.7 ± 2.3
BUSI647*	87.3 ± 2.3	85.7 ± 1.7

* Target concepts unavailable. Colors follow the paper's comparisons.



Limited Concept Supervision

BUSC-BUSBRA, Attention editor, separate ablation.



Contribution

- **BUSC benchmark.** 1,872 BUSBRA + 647 BUSI lesion cases align diagnostic labels with six clinical concepts, supporting evaluation under incomplete supervision.
- **Report-guided self-editing.** Structured reports teach residual concept corrections during training; the deployed model uses ultrasound images alone and needs no report.
- **Robust and generalizable.** Ordered risk constraints and SCMT yield 91.6% / 97.0% in-domain AUC and retain 92.72% AUC with only 20% concept-supervised samples.
- **Code: github.com/wentao-2/TRACE**

***Further improvements with an extended journal version in preparation.**

 DOI: 10.1145/3767308.3836433 **© ACM Multimedia 2026**