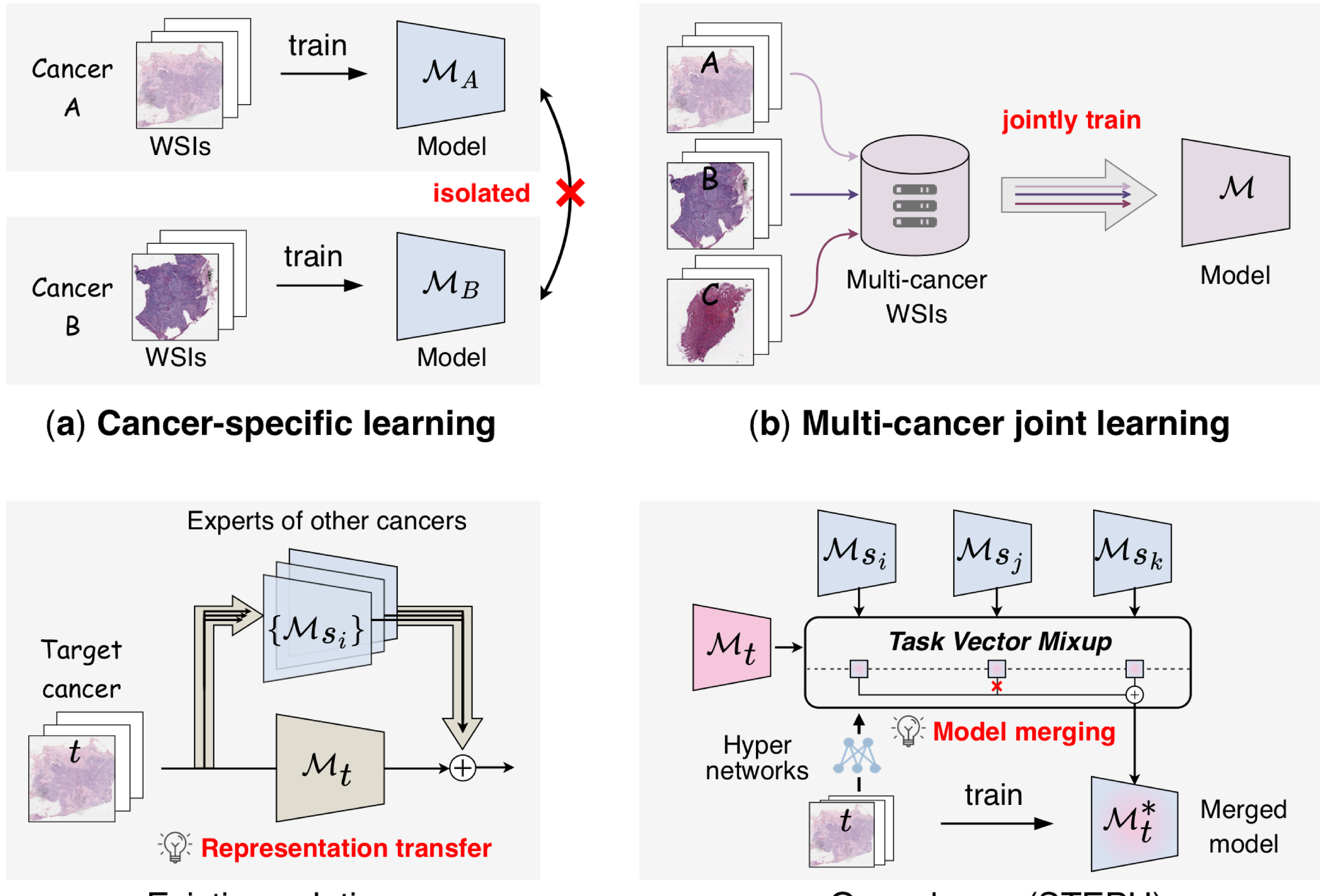
Sparse Task Vector Mixup with Hypernetworks for Efficient Knowledge Transfer in Whole-Slide Image Prognosis

Pei Liu[1], Xiangxiang Zeng[1], Tengfei Ma[1], Yucheng Xing[2], Xuanbai Ren[1], Yiping Liu[1] | [1] Hunan University [2] NUS

CVPR DENVER COLORADO JUNE 3-7, 2026

➢ Motivation

- **WSI samples are often *scarce* for one cancer type ($N \approx 1,000$), so**
  - Could WSI-based prognostic models ***generalize well***, especially when confronted with high heterogeneity in tumors?
- **Existing approaches** to addressing this problem:
  - Multi-cancer joint training: expensive computational costs
  - Representation transfer: extensive model inference



(a) **Cancer-specific learning** (b) **Multi-cancer joint learning**

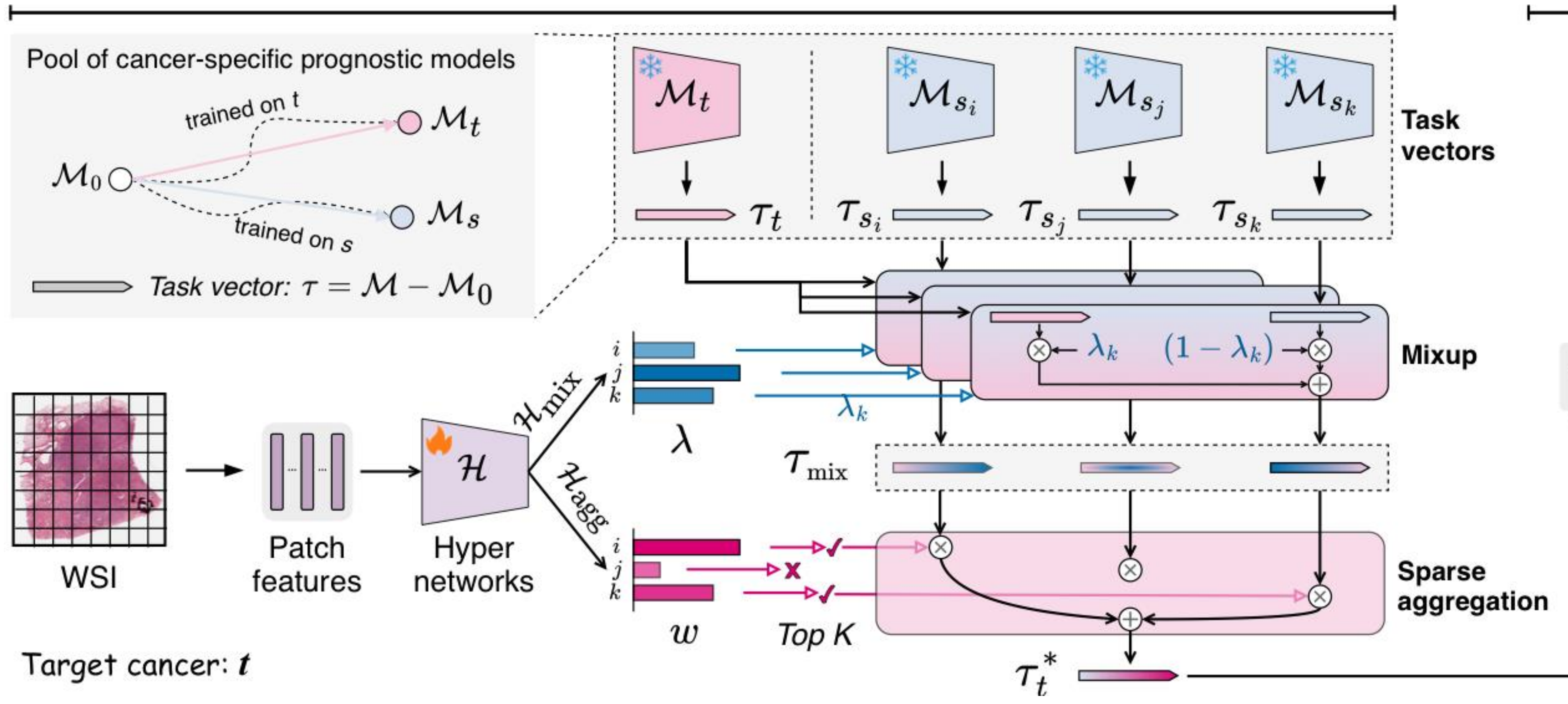
Existing solution Our scheme (STEPH)

(c) **Cross-cancer knowledge transfer**

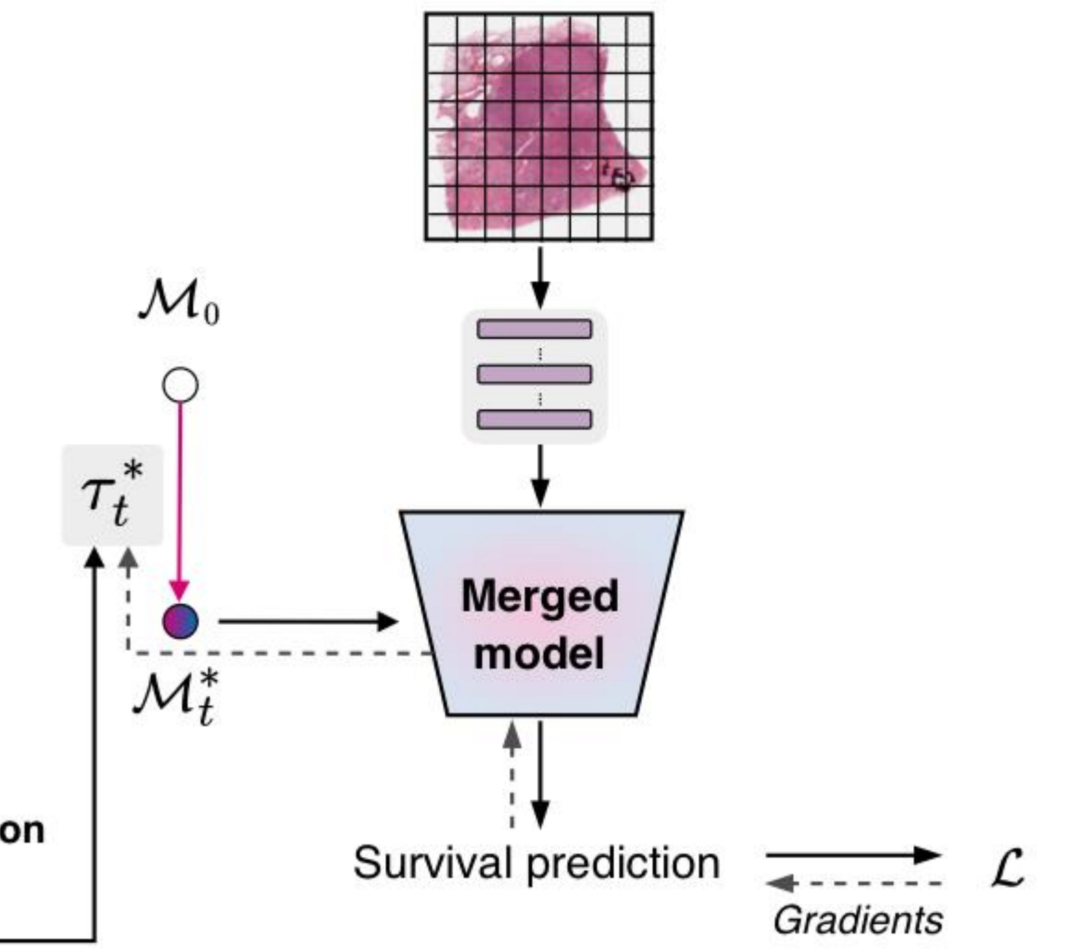
- **The proposed STEPH**:
  - A lightweight, model merging-based approach
  - Efficient cross-cancer knowledge transfer
  - Better overall performance on 13 datasets

➢ Method

STEPH (Sparse Task Vector Mixup with Hypernetworks)



Network training



Key formulas

$$\lambda = \mathcal{H}_{\text{mix}}(X)$$

$$\tau_{\text{mix}} = \lambda\tau_t + (1-\lambda)\tau_s$$

$$w = \mathcal{H}_{\text{agg}}(X)$$

$$\tau_t^* = \sum_j w_j \tau_{\text{mix},j}$$

$$\mathcal{M}_t^* = \mathcal{M}_0 + \tau_t^*$$

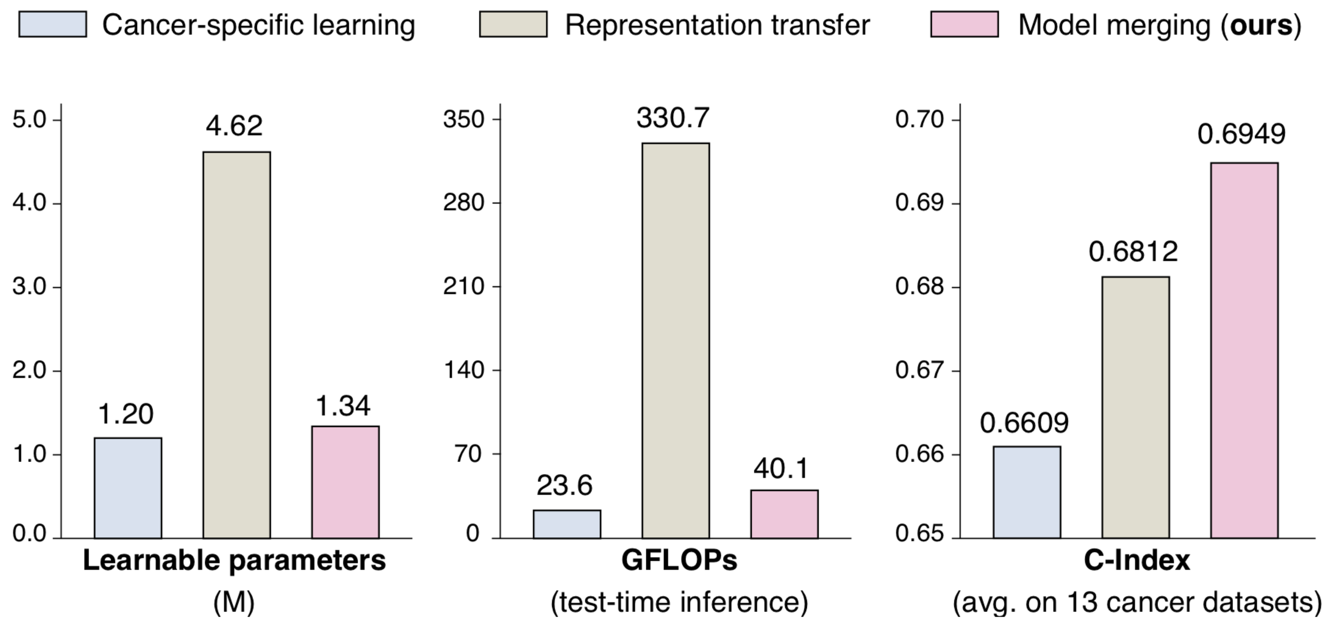
$$\mathcal{L} = \mathcal{L}_{\text{sl}} + \mathcal{L}_{\text{aux}}$$

$$\mathcal{L}_{\text{aux}} = \beta\mathcal{L}_{\text{mix}} + \gamma\mathcal{L}_{\text{agg}} = \beta\frac{\sum_j \lambda_j^2}{K} + \gamma(\log\sum_i e^{w_i})^2$$

➢ Experiments & results

Comparison with existing approaches

- improves by 5.14% and 2.01% on average
- less computational costs at inference
- performs better than existing model merging-based approaches



Ablation study

Ablation		Avg.
- On task vector mixup		
w/o mixup	fix $\lambda = 0$ (only $\{\tau_{s_i}\}$)	0.6860
	fix $\lambda = 0$ ($\tau_t \in \{\tau_{s_i}\}$)	0.6895
	fix $\lambda = 1$ (only τ_t)	0.6851
w/ mixup	trainable param. λ	0.6921
	hypernetwork-driven λ	**0.6949**
- On task vector aggregating		
w/o sparsity	hypernetwork-driven w	0.6912
w/ sparsity	trainable param. w	0.6490
	hypernetwork-driven w	**0.6949**

Visualization of task vector mixture