CROSS-MODAL HETEROGENEOUS PRIVILEGED POLICY OPTIMIZATION FOR MEDICAL IMAGE UNDERSTANDING

Anonymous authors
Paper under double-blind review

ABSTRACT

Foundation models for medical imaging now span self-supervised vision, dense supervision, and vision–language learning, offering complementary knowledge about structure, localization, and semantics. Clinical deployment, however, still favors compact vision-only models, and transferring from such heterogeneous teachers jointly is difficult because their modalities, objectives, and representation spaces do not match. We introduce HiPPO, an on-policy distillation framework that transfers heterogeneous multimodal foundation models into a compact vision-only student. Conditioned on the current student state, HiPPO samples candidate representations as binary selections over a unified codebook, scores each in the native space of every applicable teacher via teacher-specific projectors and compatibility functions, and reinforces candidates with stronger group-relative feedback. The selected representations drive position-wise visual distillation from frozen visual foundation models and privileged semantic distillation from a frozen medical vision–language model conditioned on annotation-derived views available only during training. We further establish MedBridge-Bench, spanning case-level disease diagnosis and voxel-level lesion recognition over more than 18,000 volumetric CT, MRI, and PET/CT studies. Across these settings, HiPPO improves compact medical vision models and attains competitive performance, while all teachers, privileged inputs, and policy modules are discarded at inference.

1 INTRODUCTION

Medical image understanding is rapidly entering the foundation-model era. Recent advances span self-supervised visual encoders (Oquab et al., 2024; Siméoni et al., 2025; Pérez-García et al., 2025), vision–language models (VLMs) (Zhang et al., 2023), promptable segmentation models (Kirillov et al., 2023; Ma et al., 2024a; Zhu et al., 2024), and medical vision–language models (Xu et al., 2025; Jiang et al., 2025). Trained with different architectures, objectives, modalities, and supervision signals, these models acquire complementary knowledge of anatomical structure and texture, boundary and localization cues, cross-modal semantic alignment, and diagnostic semantics. Collectively, they offer a rich but fragmented source of supervision for medical image understanding.

Despite their growing capabilities, directly deploying foundation models is often impractical due to clinical computation, privacy, and input constraints, as illustrated in Fig. 1. The central challenge is therefore to transfer this heterogeneous knowledge from diverse teachers within a unified framework while retaining a compact, vision-only student at inference. While knowledge distillation offers a natural solution (Hinton et al., 2015), jointly absorbing heterogeneous knowledge is difficult: teachers differ fundamentally in dimensionality, granularity, and representation space, making static weighting or fixed projections suboptimal. Furthermore, distillation utility varies with the student's evolving state and spatial context. Effective transfer must therefore dynamically adapt not only to what each teacher provides, but to what the student currently needs and can absorb.

A policy-based perspective naturally suits this setting: rather than committing to fixed teacher architectures, it learns which representations best serve the student's current state, allowing supervision to evolve on-policy with the learner (Ross et al., 2011; Agarwal et al., 2024; Gu et al., 2024). However, existing methods target autoregressive students with discrete token actions and response-level align-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

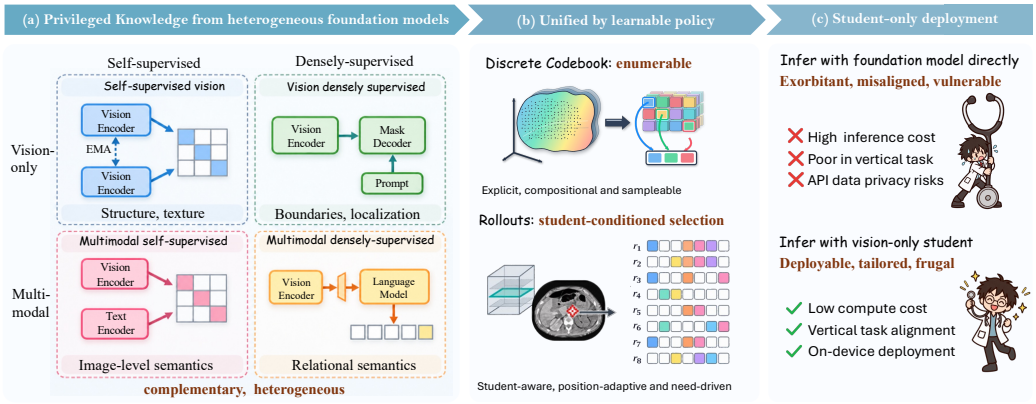

Figure 1: **Motivation of the cross-modal heterogeneous privileged distillation.** (a) Heterogeneous foundation models provide complementary privileged knowledge; (b) A discrete codebook makes their composition enumerable, dynamic rollouts make the composition adaptive; (c) Vision-only student is deployed, avoiding the cost, mismatch for domain and tasks, and potential privacy risk.

ment. In contrast, language-free medical vision students produce continuous, spatially structured representations for dense prediction, lacking natural token spaces or rollouts. Extending on-policy distillation to heterogeneous multimodal teachers therefore requires a representation-level action space, a mechanism to convert incomparable teacher outputs into feedback, and an objective that filters weak or conflicting candidates.

We address these challenges with **Cross-modal Heterogeneous Privileged Policy Optimization (HiPPO)**, framing heterogeneous multimodal knowledge transfer as student-conditioned policy optimization. Conditioned on the student state, a policy samples representations as binary selections over a unified codebook. These candidates are projected into each teacher's native space, evaluated via compatibility functions, and updated using group-relative advantages, avoiding forcing teachers into a single target space. The selected representations guide position-wise visual distillation from frozen visual foundation models and privileged semantic distillation from a medical vision–language model conditioned on training-only lesion-centered views (Pechyony & Vapnik, 2010; Lopez-Paz et al., 2015). At inference, all auxiliary modules, teachers, and privileged inputs are discarded, leaving only the vision student. We further introduce **MedBridge-Bench**, spanning 20 public datasets and three independent out-of-distribution cohorts (>18,000 volumetric CT, MRI, and PET/CT studies) to evaluate lesion detection and segmentation across modalities and distribution shifts. HiPPO improves compact medical vision models across both in-distribution and out-of-distribution settings. Our contributions are threefold:

- We formulate knowledge transfer from heterogeneous multimodal foundation models to a language-free medical vision student as a student-conditioned on-policy learning problem, where actions compose candidate representations from a shared codebook rather than select teachers.
- We propose HiPPO, which converts teacher-specific feedback into group-relative advantages for adaptive representation sampling and combines position-wise visual distillation with training-only privileged semantic supervision.
- We establish MedBridge-Bench, covering 20 public datasets, three independent out-of-distribution cohorts, and more than 18,000 volumetric studies, and demonstrate gains in case-level lesion detection and voxel-level lesion segmentation across diverse evaluation settings.

2 Related Work

Medical Foundation Models. Foundation models now encode prior knowledge at a scale far beyond any individual screening dataset, spanning 2D vision–language models (Zhang et al., 2023; Xu et al., 2025; Jiang et al., 2025), self-supervised backbones (Oquab et al., 2024; Siméoni et al., 2025), promptable segmentation models (Kirillov et al., 2023; Ma et al., 2024a; 2025), 3D self-supervised and vision–language backbones (Wu et al., 2024; Blankemeier et al., 2026; Bai et al., 2024), and

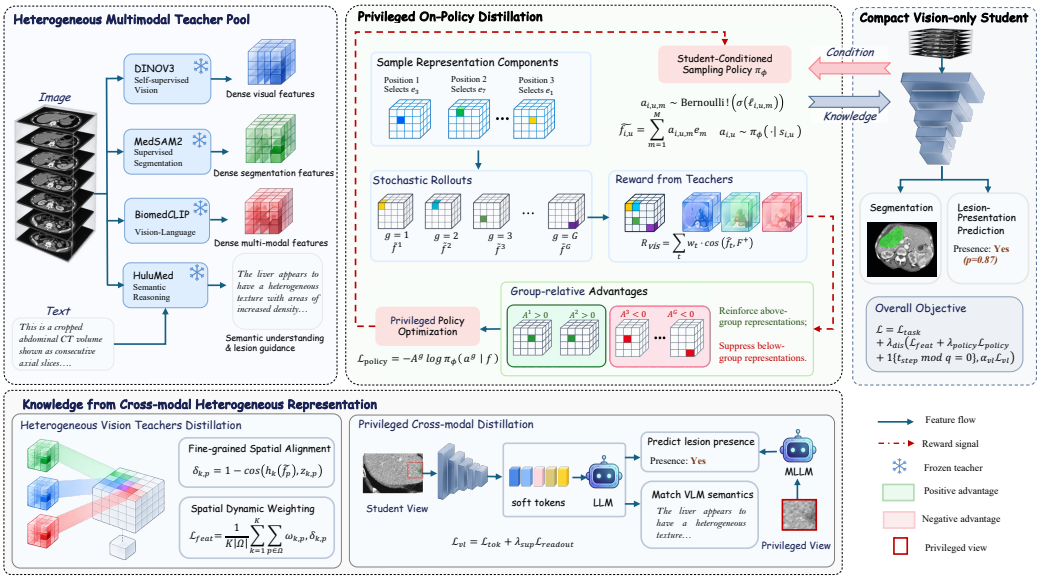

Figure 2: **Overview of HiPPO.** Frozen heterogeneous teachers give complementary feedback for advantage-guided rollout learning. A student-conditioned policy samples candidate rollouts from a unified representation codebook and scores them against the frozen visual teachers; above-group rollouts are reinforced, and the same rewards, normalized across positions, weight where visual features are distilled. A frozen medical VLM adds privileged semantic supervision via soft tokens, all distilled into one compact multi-task medical vision student.

medical multimodal LLMs (Jiang et al., 2025; Xu et al., 2025). Because they differ in architecture, objective, and supervision granularity, they encode complementary knowledge and semantics, that jointly supply priors previously unavailable at this scale.

On-Policy and Privileged Distillation. On-policy distillation (OPD) has the teacher score the student's own generations. In imitation learning this reduces the compounding error of off-policy supervision from quadratic to linear in the sequence horizon (Ross et al., 2011), motivating distillation along student-generated rather than teacher-generated sequences (Agarwal et al., 2024; Gu et al., 2024; Ko et al., 2024), which is now standard in LLM post-training (Yang et al., 2025; Zhao et al., 2026). It has recently crossed into vision–language students (Bousselham et al., 2026; Yuan et al., 2026b; Qian et al., 2026), yet supervision still flows through a text decoder. Privileged information distillation (Vapnik & Vashist, 2009; Vapnik & Izmailov, 2015) instead gives the teacher side information the student never sees; this asymmetry has been combined with OPD for VLM (Yuan et al., 2026b; Qian et al., 2026). In the dense visual setting the two asymmetries again coincide, but remain unexplored. App. E details these three lines of work and positions HiPPO against each of them.

3 Method: HiPPO

3.1 Problem Setup and Overview

Task formulation. Let $x \in \mathbb{R}^{C\times H_1\times H_2\times H_3} \in \mathcal{D}$ be a 3D medical image. A compact vision student $S_\theta = H_\theta \circ E_\theta$ encodes $f = E_\theta(x)$ and predicts $\hat{y} = H_\theta(f)$. We consider a heterogeneous multimodal teacher set indexed by $\mathcal{J} = \{1, \ldots, K\} \cup \{\text{vl}\}$:

$$\mathcal{T} = \{T_{j,\phi_j^*}\}_{j\in\mathcal{J}} = \mathcal{T}_{\text{vis}} \cup \{T_{\text{vl},\phi_{\text{vl}}^*}\}, \quad \mathcal{T}_{\text{vis}} = \{T_{k,\phi_k^*}\}_{k=1}^{K}, \quad z_j = T_{j,\phi_j^*}(v_j(x)) \in \mathcal{Z}_j, \quad (1)$$

where ϕ_j^* denotes frozen teacher parameters. These teachers are heterogeneous in modality, architecture, and representation. Input adapters v_j map the 3D volume to each teacher's input space, while learnable projectors $h_{\psi_j} : f \mapsto \mathcal{Z}_j$ align student features with the corresponding teacher representations. The two teacher groups provide asymmetric supervisory signals. Specifically, visual teachers $\mathcal{T}_{\text{vis}}$ score sampled student rollouts via dense targets from h_{ψ_k}, and the VLM T_{vl} supplies a distribution-level semantic target conditioned on projected student tokens (Sec. 3.4).

Framework overview. As shown in Fig. 2, we define the *Medical Vision On-Policy Distillation* framework to train a compact student via a student-conditioned rollout policy under frozen foundation model supervision, and instantiate it as **HiPPO (Cross-modal Heterogeneous Privileged Policy Optimization)**. Specifically, we first encode the input medical image into local student states to condition the rollout policy. The sample candidate representations from a unified codebook and score them using compatibility rewards from frozen visual teachers. Next, we compute group-relative advantages from these rewards to scale the policy gradient, and normalize them across spatial positions to dictate *where* dense features are distilled. Finally, we leverage a frozen medical VLM to supply token-level probability distributions for privileged semantic distillation. For the architecture, the student comprises a shared 3D encoder, dense segmentation decoders, and a multi-label presence head. The teacher pool includes three heterogeneous frozen visual models (DINOv3 (Siméoni et al., 2025), BiomedCLIP (Zhang et al., 2023), MedSAM2 (Ma et al., 2025)) and one frozen medical VLM (HuluMed (Jiang et al., 2025)); implementation details are deferred to App. C.1.

Training and inference. We train the overall framework by optimizing a complete objective that combines foundational task supervision with two distillation branches and the rollout term supplying their priorities. Specifically, we first define the task loss by applying a joint Dice and cross-entropy loss to the dense segmentation targets alongside a binary cross-entropy loss for multi-label presence prediction, yielding $\mathcal{L}_{\text{task}} = \mathcal{L}_{\text{dense}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}}$, where $\mathcal{L}_{\text{dense}}$ aggregates the per-target segmentation terms. We then construct the total training objective as:

$$\mathcal{L} = \mathcal{L}_{\text{task}} + \lambda_{\text{dis}}(\mathcal{L}_{\text{feat}} + \lambda_{\text{policy}}\mathcal{L}_{\text{policy}} + \mathbf{1}\{t_{\text{step}} \bmod q = 0\}\, \alpha_{\text{vl}}\mathcal{L}_{\text{vl}})\,, \tag{2}$$

where t_{step} is the optimization step and the semantic term is evaluated only every q steps to reduce the computational overhead of the frozen VLM. We detail the gradient flow and the stop-gradient mechanism applied during this periodic evaluation in App. C.5. During inference, we discard all teacher components, auxiliary projectors, and the rollout machinery. **We retain solely the compact medical vision student (E_θ, its dense decoders, and the presence head), ensuring the deployment cost remains identical to standard supervised baselines.**

3.2 Student-Conditioned Rollout Policy

For on-policy exploration across heterogeneous teacher spaces, we assemble candidate representations via discrete codebook selection rather than continuous regression. Specifically, we first condition a unified codebook on the evolving local student states to generate base representations, and then draw multiple stochastic rollouts per spatial position to construct a diverse candidate pool for immediate teacher evaluation.

Unified rollout codebook. We define a decision unit $u \in \mathcal{U}_i$ as a specific spatial index on the student's 3D bottleneck feature. We extract the student's underlying feature vector $f_{i,u}$ and apply a lightweight projection to construct the local state $s_{i,u} = \varphi(f_{i,u})$. This state construction only depends on the student parameter evolving $s_{i,u}$ without depend on teacher or supervision cues. To generate the candidate representation for position u, we utilize a unified codebook $\mathcal{E} = \{e_m\}_{m=1}^{M}$ as a discrete basis. Specifically, we predict Bernoulli logits $\ell_{i,u}$ over the codebook components via the policy network and sample binary actions to assemble the final feature:

$$a_{i,u,m} \sim \text{Bernoulli}(\sigma(\ell_{i,u,m}))\,, \qquad \tilde{f}_{i,u} = \sum_{m=1}^{M} a_{i,u,m} e_m, \qquad a_{i,u} \sim \pi_\phi(\cdot \mid s_{i,u}). \tag{3}$$

This sampled action directly constructs a candidate student representation at position u, driven by the projected student state $s_{i,u}$.

Multiple rollout candidates. For each decision unit u, we draw G independent stochastic actions from the policy $\pi_\phi(\cdot \mid s_{i,u})$ to produce G candidate representations $\{\tilde{f}_{i,u}^{(g)}\}_{g=1}^{G}$, termed *rollouts*. Each rollout is an on-policy representation constructed solely from the projected student state $s_{i,u}$ and evaluated immediately via teacher feedback. Since each action vector activates distinct codebook components, these G rollouts provide diverse representation hypotheses for the same spatial position.

3.3 ADVANTAGE-GUIDED ROLLOUT LEARNING

To reinforce candidate representations that align with heterogeneous visual teachers, we compute group-relative advantages from compatibility rewards and utilize them to scale the policy gradients.

Visual teacher rewards. Let $\emptyset \neq \mathcal{K}_{i,u} \subseteq \{1, \dots, K\}$ index the visual teachers providing valid guidance for decision unit u. We map each rollout representation $\tilde{f}_{i,u}^{(g)}$ via a teacher-specific projector h_k and evaluate it against the teacher target $z_{k,i,u}$ using a compatibility function q_k, yielding the rollout reward:

$$r_{i,u}^{(g)} = \frac{1}{|\mathcal{K}_{i,u}|} \sum_{k \in \mathcal{K}_{i,u}} q_k\Big(h_k(\tilde{f}_{i,u}^{(g)}), z_{k,i,u}\Big), \tag{4}$$

where q_k denotes the cosine agreement defined in Eq. 7. Since a view-level scalar cannot guide dense spatial rollouts, we exclude the medical VLM from this reward. We therefore reformulate its supervision as a differentiable distributional target (Eq. 11). Agreement is therefore evaluated exclusively on representations, as frozen encoders provide no mask-level references. We apply this identical reward formulation to derive the policy advantage by normalizing across rollouts within a position, and to determine the distillation difficulty in Sec. 3.4 by normalizing across positions within a batch.

Group-relative advantage. Given G candidate rollouts for the same state $s_{i,u}$, we compute the group-relative advantage (Shao et al., 2024):

$$A_{i,u}^{(g)} = (r_{i,u}^{(g)} - \bar{r}_{i,u})/(\sqrt{G^{-1} \sum_{g'=1}^{G} \Big(r_{i,u}^{(g')} - \bar{r}_{i,u}\Big)^2} + \epsilon), \qquad \bar{r}_{i,u} = \frac{1}{G} \sum_{g=1}^{G} r_{i,u}^{(g)}, \tag{5}$$

where $\epsilon > 0$ ensures numerical stability. This advantage quantifies whether a specific rollout achieves above- or below-group visual-teacher consistency conditioned strictly on the projected student state $s_{i,u}$.

Policy optimization. We optimize the policy by minimizing a joint objective that combines a differentiable alignment loss, a group-relative policy gradient, and an entropy regularization term:

$$\mathcal{L}_{\text{policy}} = \frac{1}{|\mathcal{K}|\, G} \sum_{k \in \mathcal{K}} \sum_{g=1}^{G} \Big[1 - q_k\big(h_k(\tilde{f}^{(g)}), z_k\big)\Big] - \frac{1}{G} \sum_{g=1}^{G} \text{sg}\Big(A^{(g)}\Big) \log \pi_\phi\Big(a^{(g)} \mid s\Big) - \beta\, \mathcal{H}[\pi_\phi(\cdot \mid s)], \tag{6}$$

where $\text{sg}(\cdot)$ blocks gradient flow through the advantage, β controls the entropy bonus to prevent action collapse, and $\mathcal{K} = \mathcal{K}_{i,u}$ denotes the valid-teacher set from Eq. 4 (subscripts omitted for brevity). Since the sampled codebook composition enters the advantage strictly as a detached scalar, the alignment term provides the sole differentiable path shaping the codebook representations, while the policy gradient exclusively governs discrete component selection.

3.4 PRIORITIZED VISUAL DISTILLATION

2.5D-to-3D alignment. Each visual teacher processes pseudo-RGB axial views constructed from neighboring slices. We extract its 2D token grids in-plane at the bottleneck positions and aggregate them within student depth bins to form aligned targets $z_{k,p} \in \mathbb{R}^{C_k}$ for spatial positions $p \in \Omega$. A projector h_k maps the deterministic student representation f_p into the feature space of teacher k. We measure their agreement and the induced distance as:

$$q_k(u, v) = \cos(u, v), \qquad \delta_{k,p} = 1 - q_k(h_k(f_p),\, z_{k,p})\,. \tag{7}$$

The distance metric $\delta_{k,p}$ directly distills the deterministic student representation f_p. The sampled rollouts influence this alignment objective exclusively through the per-position weighting defined below. App. C.2 details the view construction and axial aggregation.

Prioritized weights. We define the per-position distillation weight using three detached criteria. First, the alignment-difficulty criterion reuses the rollout reward from Eq. 4, normalizing it across spatial positions rather than across rollouts:

$$\bar{q}_{k,p} = \frac{1}{G} \sum_{g=1}^{G} q_k\Big(h_k(\tilde{f}_p^{(g)}), z_{k,p}\Big), \qquad D_{k,p} = \Big[\, t - \text{clip}\big(\tfrac{\bar{q}_{k,p} - \mu_k}{\sigma_k + \epsilon}, -b, b\big)\Big]_+, \tag{8}$$

where μ_k and σ_k denote the batch mean and standard deviation of $\bar{q}_{k,\cdot}$ over Ω, t specifies the target agreement level, and b bounds the influence of batch outliers. Second, the in-plane transition salience exploits the axes resolved independently by the teachers:

$$\bar{G}_p = \frac{1}{2C}\sum_{c=1}^{C}(|\Delta_h f_c|_p + |\Delta_w f_c|_p)\,, \qquad G_p = \frac{\bar{G}_p - \min_{q\in\Omega_n}\bar{G}_q}{\max_{q\in\Omega_n}\bar{G}_q - \min_{q\in\Omega_n}\bar{G}_q + \epsilon} \in [0,1], \quad (9)$$

where Δ_h and Δ_w compute forward differences under replicate padding. Third, the segmentation-error criterion flags currently misclassified bottleneck positions via $\eta_p = \mathbf{1}\{\hat{y}_p \neq y_p\}$, evaluated against the training label (App. C.2). We combine these three criteria using an element-wise maximum and normalize the result to unit mean:

$$u_{k,p} = \max\left(G_p,\ D_{k,p},\ \eta_p\right), \quad \omega_{k,p} = \frac{u_{k,p}}{|\Omega|^{-1}\sum_{q\in\Omega} u_{k,q} + \epsilon}, \quad \mathcal{L}_{\text{feat}} = \frac{1}{|\mathcal{K}|\,|\Omega|}\sum_{k\in\mathcal{K}}\sum_{p\in\Omega}\omega_{k,p}\,\delta_{k,p}. \quad (10)$$

Here, $D_{k,p}$ upweights positions where rollouts lag behind the teacher relative to the batch, while G_p and η_p retain structurally salient or misclassified regions. The hinge operation prevents well-aligned positions from gaining emphasis. App. C.2 provides the derivation, tail-risk interpretation, and visualizations of the realized weights (Fig. 11).

Privileged semantic matching. Dense visual teachers supply spatial targets but lack diagnostic semantics. We therefore introduce a frozen medical VLM supervised under privileged information (Pechyony & Vapnik, 2010; Lopez-Paz et al., 2015). Specifically, the annotation defines a magnified lesion-centered view visible exclusively to the teacher. A trainable projection Ψ maps the student's bottleneck features into N_s soft tokens within the VLM's embedding space. We condition the VLM jointly on these tokens and the student's own crop, forcing any diagnostic semantics to originate strictly from the student's encoded representations. This semantic path remains entirely disjoint from the rollout codebook (Sec. 3.2) since it involves neither discrete actions nor RL rewards, relying solely on a differentiable divergence. Let P_τ and Q_τ denote the teacher and student-conditioned distributions at generated position τ. We restrict both distributions to the teacher's Top-K_{tok} tokens and renormalize them. The token-matching term aligns these restricted distributions via Jensen–Shannon divergence. Simultaneously, the readout term queries the conditioned VLM using per-class yes/no prompts under view-level labels to anchor the token alignment to the clinical annotation:

$$\mathcal{L}_{\text{vl}} = \mathcal{L}_{\text{tok}} + \lambda_{\text{sup}}\mathcal{L}_{\text{readout}} = T^{-1}\textstyle\sum_\tau \text{JSD}(P_\tau \,\|\, Q_\tau) + \lambda_{\text{sup}}\,\text{BCE}(\hat{\mathbf{c}}_{\text{VLM}}, \mathbf{c})\,. \quad (11)$$

This Top-K_{tok} restriction ensures numerical stability of the JSD computation even when teacher-preferred tokens carry little student mass, and unlike the freely learned presence head the readout forces view-level semantics into the frozen VLM's space. App. C.4 details view construction, lesion-free handling and the Top-K_{tok} restriction.

4 Experiment

4.1 Benchmark: MedBridge-Bench

We construct *MedBridge-Bench* from public and private datasets (Soler et al., 2010; Akin et al., 2016; Linehan et al., 2016; National Cancer Institute Clinical Proteomic Tumor Analysis Consortium, 2018; Hernandez Petzsche et al., 2022; Gatidis et al., 2024; Bartnik et al., 2024; Jackson et al., 2025; Peeters et al., 2025; Luo et al., 2025; Bassi et al., 2025; Wu et al., 2026; Li et al., 2026b), comprising 20,689 scored instances drawn from 18,595 unique volumetric studies across CT, MRI, and PET/CT. The benchmark is organized into five sub-benchmarks: *MedBridge-CT.TumorExt* is a CT tumor dataset with an internal, in-distribution (IID) test set and external, out-of-distribution (OOD) cohorts, enabling assessment of generalization under distribution shift. *MedBridge-CT.TumorInt* is a second CT tumor dataset with internal evaluation only. *MedBridge-CT.Nodule* is a CT dataset for pulmonary nodule analysis. *MedBridge-MR* contains MRI datasets for lesion analysis. *MedBridge-PET/CT* contains PET/CT datasets for lesion analysis. Across these sub-benchmarks, MedBridge-Bench supports lesion detection/classification, dense voxel-level segmentation, and out-of-distribution generalization assessment. Dataset compositions, splits, and preprocessing protocols are detailed in App. A.

Table 1: Lesion-level classification and segmentation results on the internal test set of MedBridge-CT.TumorExt.{organ}. Models in the *w/ LLMs* group were fine-tuned on the training set as in Bassi et al. (2025); BAcc is balanced accuracy and AUC the area under the receiver-operating-characteristic curve. Best results are highlighted in **bold** and runners-up underlined.

Model	Liver Tumor (%)				Kidney Tumor(%)				Pancreas Tumor(%)				BAcc	AUC	F1	Dice
	BAcc	AUC	F1	Dice	BAcc	AUC	F1	Dice	BAcc	AUC	F1	Dice				
w/o LLMs																
nnU-Net (Isensee et al., 2021)	77.10	82.66	49.14	44.41	69.68	73.27	50.99	43.55	51.23	45.02	21.65	17.45	66.00	66.98	40.60	35.14
SwinUNETR (Hatamizadeh et al., 2021)	82.29	85.71	56.19	53.67	74.76	77.21	57.89	55.63	64.86	68.67	29.57	24.75	73.97	77.20	47.89	44.68
3D UX-Net (Lee et al., 2022)	83.09	85.47	53.76	55.90	75.49	77.25	58.19	58.92	66.90	69.38	33.25	25.56	75.16	77.37	48.40	46.79
CLIP-driven (Liu et al., 2023)	82.75	86.39	55.26	47.21	75.41	77.92	57.84	52.38	62.90	65.93	29.01	21.75	73.69	76.75	47.37	40.45
TransUNet (Chen et al., 2024)	75.58	76.35	52.83	31.02	68.03	68.84	51.55	23.52	55.41	55.05	21.69	3.05	66.34	66.75	42.02	19.20
UniMiSS+ (Xie et al., 2024)	84.97	85.18	**58.73**	58.42	78.45	77.72	61.77	**62.44**	76.25	79.25	47.92	24.50	79.89	80.72	56.14	48.45
U-Mamba (Ma et al., 2024b)	79.33	84.74	50.47	53.58	75.89	77.28	57.95	56.44	61.78	64.76	27.96	23.88	72.33	75.59	45.46	44.63
SuPreM (Li et al., 2024)	83.77	86.50	56.04	58.57	78.36	79.83	63.06	61.48	71.57	75.12	37.35	29.05	77.90	80.48	52.15	49.70
GF-Screen (Wu et al., 2026)	73.26	73.77	54.69	28.67	74.90	75.32	63.10	40.18	63.15	63.05	40.54	11.72	70.44	70.71	52.78	26.86
w/ LLMs																
CT2Rep (Hamamci et al., 2024)	56.74	–	26.70	–	54.08	–	32.21	–	46.25	–	0.00	–	52.36	–	19.64	–
M3D (Bai et al., 2024)	43.47	–	14.11	–	33.68	–	10.51	–	32.57	–	5.87	–	36.57	–	10.16	–
RadFM (Wu et al., 2025)	49.56	–	7.02	–	50.38	–	7.91	–	49.94	–	0.00	–	49.96	–	4.98	–
RadGPT (Bassi et al., 2025)	67.57	–	33.74	–	62.21	–	31.92	–	**85.46**	–	44.80	–	71.75	–	36.82	–
CT-Chat (Hamamci et al., 2026)	49.54	–	6.36	–	53.03	–	29.48	–	58.40	–	25.08	–	53.65	–	20.31	–
Merlin (Blankemeier et al., 2026)	66.06	–	44.97	–	61.45	–	40.32	–	59.52	–	26.51	–	62.34	–	37.27	–
HiPPO (ours)	**85.11**	**91.67**	57.01	**61.20**	**82.28**	**89.80**	**68.86**	60.38	83.23	**89.53**	**58.00**	**35.64**	**83.54**	**90.33**	**61.29**	**52.41**

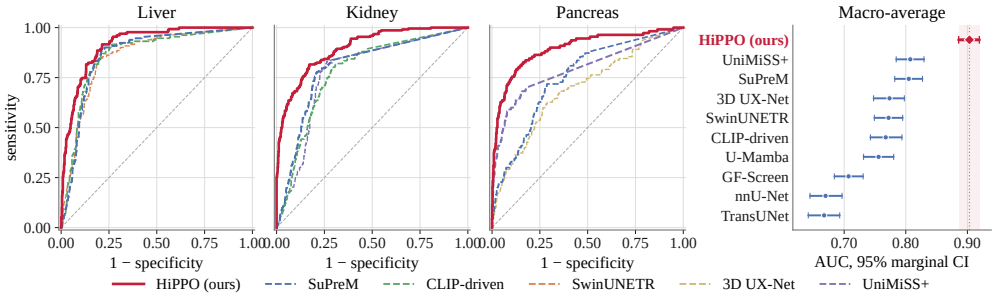

Figure 3: Lesion-level ROC curves per organ and macro-averaged AUC with each model's own 95% case-level bootstrap confidence interval (CI) on the internal test set of MedBridge-CT.TumorExt. All nine baseline intervals are disjoint from HiPPO's at the macro level.

Table 2: Lesion classification and segmentation performance on the out-of-distribution test set of MedBridge-CT.TumorExt.{organ}. Best results are shown in **bold** and runners-up underlined.

Model	Liver Tumor (%)				Kidney Tumor (%)			Pancreas Tumor (%)			
	BAcc	AUC	F1	Dice	BAcc	AUC	F1	BAcc	AUC	F1	Dice
nnU-Net (Isensee et al., 2021)	67.88	77.22	65.74	29.21	70.39	78.14	86.10	51.84	48.03	40.27	18.83
SwinUNETR (Hatamizadeh et al., 2021)	68.89	81.49	62.47	28.01	78.17	85.01	83.54	63.59	67.95	45.95	23.14
3D UX-Net (Lee et al., 2022)	74.31	85.82	74.20	39.50	82.49	87.52	84.88	66.41	72.03	49.48	25.47
CLIP-driven (Liu et al., 2023)	61.27	64.41	62.48	11.21	57.36	73.90	84.29	52.07	51.31	35.74	4.28
TransUNet (Chen et al., 2024)	65.29	65.69	53.45	12.87	78.14	80.11	77.34	54.88	57.50	24.22	4.27
UniMiSS+ (Xie et al., 2024)	72.89	80.93	70.99	29.82	88.57	89.61	89.75	69.60	71.58	54.60	21.45
U-Mamba (Ma et al., 2024b)	65.07	82.61	52.27	33.60	87.56	91.38	89.82	61.22	71.39	37.94	22.79
SuPreM (Li et al., 2024)	72.82	82.57	71.38	31.06	80.29	89.73	78.83	66.07	70.49	49.03	26.45
GF-Screen (Wu et al., 2026)	65.00	64.75	47.70	11.98	85.55	86.20	83.11	59.32	59.60	31.86	8.05
HiPPO (ours)	**84.34**	**91.31**	**89.92**	**47.07**	**90.52**	**92.48**	**92.04**	**77.56**	**86.50**	**67.04**	**31.68**

4.2 Comparative Results

HiPPO delivers consistently strong results across lesion detection, segmentation, malignancy screening, and external evaluation. On *MedBridge-CT.TumorExt*, it ranks first in macro BAcc, AUC, F1, and Dice, reaching a macro AUC of 90.33% and a macro Dice of 52.41% (Tab. 1); the improvements are supported by paired bootstrap intervals and multiplicity-corrected DeLong tests (Figs. 3 and 5). This advantage persists on the external liver, pancreas, and kidney cohorts, where thresholds calibrated on the IID cohort are kept fixed and HiPPO ranks first on each cohort, with AUCs of

86.50%–92.48% (Tab. 2). It also achieves the best macro results in six-class malignancy screening, including a macro AUC of 89.33% (Tab. 3). Beyond abdominal CT, we retain the same architecture and distillation framework while adopting dataset-specific training settings for stroke MRI, PET/CT, and thoracic CT. HiPPO leads the reported stroke-MRI metrics, the thoracic-CT nodule metrics, and the PET/CT detection results (Tab. 4). **Together, these results show that its gains are not confined to a single organ, imaging protocol, or modality.** Besides, a shared head-free linear probe evaluates whether the features themselves encode lesion evidence (Fig. 4). All frozen encoders are evaluated with the same lesion-versus-background readout, so the resulting *probe AUC* measures representation separability rather than downstream detection performance. The teachers achieve probe AUCs of 0.886–0.930, **confirming that their supervision is lesion-related.** Under the same training schedule, distillation raises the student's probe AUC from 0.898 to 0.922 at the probed encoder stage, bringing its representation into the teacher range. **This result shows that distillation improves lesion–background separability directly at the representation level.**

Table 3: Lesion-level classification results on the MedBridge-CT.TumorInt{organ}. Group-wise scores are arithmetic means over member classes; macro metrics average all 6 classes (%). $\text{Se}_{@95}$: sensitivity at specificity $\geq 95\%$; $\text{Sp}_{@95}$: specificity at sensitivity $\geq 95\%$.

Model	Lung (%)			Abdominal Solid (%)			Abdominal Hollow (%)			AUC	$\text{Se}_{@95}$	$\text{Sp}_{@95}$
	AUC	$\text{Se}_{@95}$	$\text{Sp}_{@95}$	AUC	$\text{Se}_{@95}$	$\text{Sp}_{@95}$	AUC	$\text{Se}_{@95}$	$\text{Sp}_{@95}$			
nnU-Net (Isensee et al., 2021)	89.35	49.61	54.78	79.20	15.78	43.31	65.19	4.35	33.76	78.56	19.51	43.63
SwinUNETR (Hatamizadeh et al., 2021)	74.60	0.00	35.67	54.55	0.00	19.11	30.60	0.00	5.10	53.90	0.00	19.53
3D UX-Net (Lee et al., 2022)	81.75	0.00	53.50	62.14	0.00	31.37	36.61	0.00	14.65	61.15	0.00	32.27
CLIP-driven (Liu et al., 2023)	83.38	32.56	56.69	66.86	5.38	21.97	46.16	0.00	11.46	66.16	9.01	26.01
TransUNet (Chen et al., 2024)	92.81	55.81	68.79	89.17	**58.39**	51.43	63.89	8.70	26.11	85.56	49.68	50.11
UniMiSS+ (Xie et al., 2024)	69.50	0.00	33.76	52.67	0.00	25.00	28.99	0.00	3.82	51.53	0.00	22.93
VoCo (Wu et al., 2024)	80.17	20.16	42.04	60.19	4.72	21.18	32.35	0.00	3.82	58.88	6.50	21.76
U-Mamba (Ma et al., 2024b)	83.09	37.21	52.87	67.06	3.91	23.89	47.83	0.00	29.94	66.53	8.81	29.72
SuPreM (Li et al., 2024)	74.41	0.00	44.59	55.76	0.00	24.04	34.31	0.00	7.01	55.29	0.00	24.63
GF-Screen (Wu et al., 2026)	93.73	28.68	**79.62**	86.45	23.06	45.22	83.74	0.00	**70.70**	87.21	20.15	55.20
HiPPO (ours)	**94.08**	**69.77**	74.52	**89.48**	52.68	**61.46**	**84.01**	**21.74**	69.43	**89.33**	**50.37**	**64.97**

Table 4: Results on the **MedBridge-CT.Nodule**, **MedBridge-MR** and **MedBridge-PET/CT** subbenchmarks, the last as separate segmentation (**-Seg**) and classification (**-Cls**) tasks; **M.B.** abbreviates **MedBridge** in the header.

Model	M.B.-CT.Nodule			M.B.-MR			M.B.-PET/CT-Seg			M.B.-PET/CT-Cls			
	BAcc	AUC	F1	Dice	NSD	HD95 (↓)	Dice	NSD	HD95 (↓)	Dice	BAcc	AUC	F1
nnU-Net (Isensee et al., 2021)	75.10	82.30	32.65	77.70	85.44	19.23	84.59	68.40	13.35	64.25	82.04	86.82	86.76
SwinUNETR (Hatamizadeh et al., 2021)	71.31	79.13	29.72	74.17	82.96	8.86	83.58	65.68	12.26	62.08	79.31	83.41	84.44
3D UX-Net (Lee et al., 2022)	74.64	82.01	33.77	75.92	84.40	9.20	84.76	65.63	11.20	62.59	68.80	69.41	79.53
CLIP-driven (Liu et al., 2023)	67.19	71.98	25.20	76.53	85.06	6.38	82.42	63.02	19.33	61.92	78.89	83.98	82.17
TransUNet (Chen et al., 2024)	76.42	83.41	37.91	74.95	82.23	27.82	77.71	53.01	16.94	61.69	85.75	91.48	88.38
UniMiSS+ (Xie et al., 2024)	75.83	82.08	42.42	76.13	85.33	5.81	83.65	62.79	16.17	61.52	68.14	68.00	82.91
VoCo (Wu et al., 2024)	68.48	74.41	30.41	76.20	84.77	10.14	84.86	66.08	23.32	64.56	74.19	80.14	81.64
U-Mamba (Ma et al., 2024b)	78.27	80.68	39.57	75.03	84.51	7.30	83.68	65.00	13.62	61.31	73.00	76.74	79.30
SuPreM (Li et al., 2024)	73.06	76.85	29.46	75.75	84.00	15.29	84.18	65.74	**11.13**	63.92	75.53	77.51	84.53
HiPPO (ours)	**80.22**	**85.61**	**43.86**	**79.00**	**87.15**	**5.69**	**85.21**	**68.49**	12.44	**65.97**	**88.69**	**93.78**	**91.67**

4.3 Ablation Studies and Analysis

Tab. 5 evaluates HiPPO variants under identical data, student architecture, and training schedule. Panel (a) confirms both the strength and complementarity of its supervision sources: each single-expert variant substantially outperforms the baseline, while combining all three visual experts with privileged semantic supervision achieves the best AUC, BAcc, and F1; removing any source consistently reduces all three metrics. Panel (b) further isolates the advantage of the proposed distillation strategy. The student-conditioned rollout attains higher AUC and BAcc than deterministic projection, soft mixture, input-independent dropout, and uniform random rollout, showing that the gain comes from learning discrete, sample-adaptive expert compositions rather than from projection, stochasticity, or feature mixing alone. Removing codebook-GRPO lowers all metrics and causes a particularly large F1 drop of 4.11 points, demonstrating the importance of optimizing the rollout policy with teacher-relative rewards. Removing weight unification produces the largest joint decline in AUC and BAcc, confirming that feeding rollout-derived difficulty back into prioritized feature distillation is essential for converting the learned compositions into effective student supervision.

Table 5: Mechanism ablations on *MedBridge-CT.TumorExt* (performances are macro-averaged over three lesion classes). (a) varies the teacher pool; (b) varies the rollout policy and the number of codebooks R sharing one reward definition, each of size M and read by its own gate head.

(a) Teacher pool (student always included)

Variant	Bio.	DINO	MedS.	Priv.	AUC	BAcc	F1
Baseline	-	-	-	-	66.00	66.98	40.60
HiPPO-S	✓				88.88	82.00	57.62
		✓			88.57	81.28	59.60
			✓		89.30	82.96	61.01
HiPPO-M	✓		✓	✓	87.92	78.52	56.44
		✓	✓	✓	89.28	81.38	60.17
	✓	✓		✓	87.61	78.74	56.45
	✓	✓	✓		88.77	82.27	59.17
HiPPO-M (full version)	✓	✓	✓	✓	**90.33**	**83.54**	**61.29**

(b) Rollout policy and codebook

Variant	AUC	BAcc	F1
Deterministic projection	89.69	83.06	57.59
Soft mixture	88.99	82.36	61.59
Input-independent dropout	89.25	82.46	57.48
Uniform random rollout	88.69	82.15	58.15
w/o codebook-GRPO term	89.49	82.88	57.18
w/o weight unification	88.82	81.56	59.25
Codebooks R (K=3 experts, M=64 fixed)			
R=2	89.34	82.91	**61.92**
R=3	83.04	75.17	51.01
HiPPO (ours), R=1	**90.33**	**83.54**	61.29

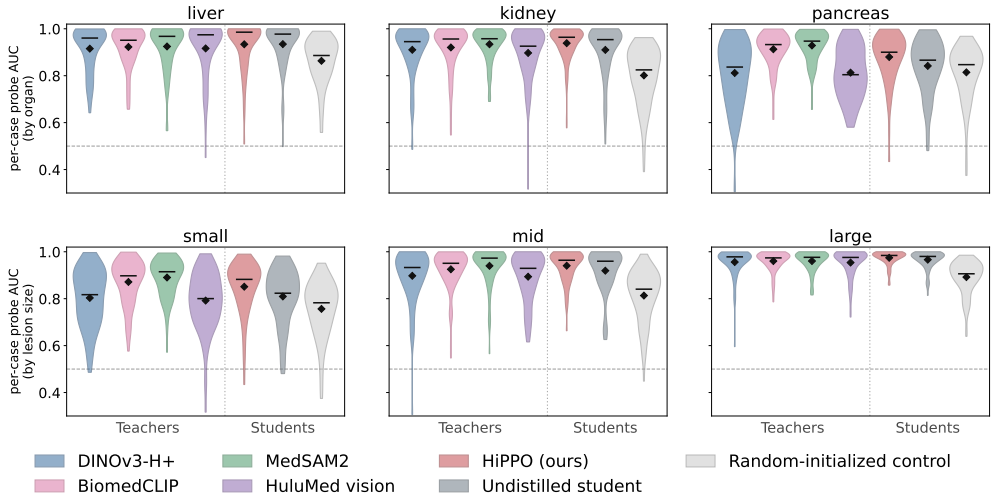

Figure 4: **Per-case lesion–background separability** of four frozen teachers and three student arms under a shared 5-fold cross-validated linear probe on projected 14×14 features at encoder stage *s3*. The teachers provide separable lesion-related features, while the distilled student surpasses its undistilled counterpart and approaches the teacher range, confirming feature-level transfer.

A single codebook provides the strongest overall result: R=2 yields only a small F1 increase at the expense of AUC and BAcc, whereas R=3 degrades all metrics. **Together, these results show that HiPPO benefits not only from heterogeneous teachers, but specifically from their student-conditioned composition and the unified transfer of rollout feedback into feature learning.**

5 Conclusion

We presented HiPPO, which distills cross- modal heterogeneous frozen foundation models into a compact 3D medical vision student. A student-conditioned policy uses visual-teacher feedback to select candidate representations and prioritize dense feature distillation, while a frozen medical VLM provides privileged semantic supervision from lesion-centered views available only during training. We also established *MedBridge-Bench*, integrating 20 public datasets and three external cohorts over more than 18,000 volu- metric CT, MRI, and PET/CT studies for case-level lesion detection and voxel-level segmentation. Across these diverse settings, HiPPO achieves consistently strong performance, with particularly clear advantages under distribution shift. All teachers, privileged inputs, and policy modules are removed at inference, leaving only the compact student. The current formulation is tailored to dense 3D medical vision and has been evaluated with a fixed teacher pool; extending it to broader teacher families and dimension-adaptive transfer across 2D medical images and 4D spatiotemporal medical data remains future work.

AI USE DISCLOSURE

In this work, we used generative AI tools to assist with translation, code implementation and debugging, feedback on the presentation of research methodology, and language-level refinement of author-developed mathematical arguments and interpretations of author-provided experimental results. We did not use generative AI to generate synthetic datasets, independently develop theoretical or conceptual frameworks, supply critical ingredients for proofs, propose or refine hypotheses, design experiments, clean or reformat datasets, or conduct qualitative or thematic data analysis; the remaining required-disclosure tasks were not applicable. Additionally, we used generative AI for manuscript drafting and readability editing, and formatting tables and mathematical expressions. We reviewed all AI-assisted work: mathematical statements and result descriptions were checked by the authors, and AI-assisted code was manually inspected and tested. The tools did not generate experimental data or independently determine the conclusions of the study. We take responsibility for the final content of this work, including all text, claims, code, and artifacts produced.

ETHICS STATEMENT

This study was conducted in strict accordance with ethical standards and applicable privacy regulations. Our dataset comprises 175 private clinical cases. The retrospective collection and use of this data were reviewed and approved by the relevant Institutional Review Board (IRB) / Ethics Committee, and the requirement for informed consent was waived given the retrospective nature of the study. All patient records were rigorously de-identified and anonymized prior to analysis to ensure compliance with relevant data protection laws, posing no foreseeable risk of harm or privacy breach. Furthermore, we have implemented systematic protocols to mitigate potential biases in both annotations and model evaluation. We disclose no conflicts of interest or sponsorship that could influence the results reported in this paper.

REPRODUCIBILITY STATEMENT

We have already elaborated on the method proposed, experimental configurations, and benchmarks used in the experiments in this paper. Furthermore, we declare that the code used in this work will be released.

REFERENCES

H. J. W. L. Aerts, L. Wee, E. Rios Velazquez, R. T. H. Leijenaar, C. Parmar, P. Grossmann, S. Carvalho, J. Bussink, R. Monshouwer, B. Haibe-Kains, D. Rietveld, F. Hoebers, M. M. Rietbergen, C. R. Leemans, A. Dekker, J. Quackenbush, R. J. Gillies, and P. Lambin. Data from NSCLC-Radiomics (version 4). The Cancer Imaging Archive, 2014. URL https://doi.org/10.7937/K9/TCIA.2015.PF0M9REI.

Rishabh Agarwal, Nino Vieillard, Yongchao Zhou, Piotr Stanczyk, Sabela Ramos Garea, Matthieu Geist, and Olivier Bachem. On-policy distillation of language models: Learning from self-generated mistakes. In *International Conference on Learning Representations*, volume 2024, pp. 21246–21263, 2024.

O. Akin, P. Elnajjar, M. Heller, R. Jarosz, B. J. Erickson, S. Kirk, Y. Lee, M. W. Linehan, R. Gautam, R. Vikram, K. M. Garcia, C. Roche, E. Bonaccio, and J. Filippini. The cancer genome atlas kidney renal clear cell carcinoma collection (TCGA-KIRC) (version 3) [data set]. The Cancer Imaging Archive, 2016. URL https://doi.org/10.7937/K9/TCIA.2016.V6PBVTDR.

Natália Alves, Megan Schuurmans, Derya Yakar, Pierpaolo Vendittelli, Geert Litjens, John Hermans, and Henkjan Huisman. The panorama challenge: Public training and development dataset (1). Zenodo, April 2024. URL https://doi.org/10.5281/zenodo.11034178.

Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Annette Kopp-Schneider, Bennett A Landman, Geert Litjens, Bjoern Menze, Olaf Ronneberger, Ronald M Summers, et al. The medical segmentation decathlon. *Nature communications*, 13(1):4128, 2022.

Fan Bai, Yuxin Du, Tiejun Huang, Max Q-H Meng, and Bo Zhao. M3d: Advancing 3d medical image analysis with multi-modal large language models. *arXiv preprint arXiv:2404.00578*, 2024.

S. Bakr, O. Gevaert, S. Echegaray, K. Ayers, M. Zhou, M. Shafiq, H. Zheng, W. Zhang, A. Leung, M. Kadoch, J. Shrager, A. Quon, D. Rubin, S. Plevritis, and S. Napel. Data for NSCLC Radiogenomics (version 4). The Cancer Imaging Archive, 2017. URL https://doi.org/10.7937/K9/TCIA.2017.7HS46ERV.

Krzysztof Bartnik, Tomasz Bartczak, Mateusz Krzyziński, Krzysztof Korzeniowski, Krzysztof Lamparski, Piotr Węgrzyn, Eric Lam, Mateusz Bartkowiak, Tadeusz Wróblewski, Katarzyna Mech, et al. Waw-tace: a hepatocellular carcinoma multiphase ct dataset with segmentations, radiomics features, and clinical data. *Radiology: Artificial Intelligence*, 6(6):e240296, 2024.

Pedro RAS Bassi, Mehmet Can Yavuz, Ibrahim Ethem Hamamci, Sezgin Er, Xiaoxi Chen, Wenxuan Li, Bjoern Menze, Sergio Decherchi, Andrea Cavalli, Kang Wang, et al. Radgpt: Constructing 3d image-text tumor datasets. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 23720–23730, 2025.

Jinhe Bi, Peng Liao, Zengjie Jin, Volker Tresp, Fei Shen, Yunpu Ma, Tat-Seng Chua, et al. Opd-v: Visual on-policy self-distillation with modality balance. *arXiv preprint arXiv:2608.05131*, 2026.

Louis Blankemeier, Ashwin Kumar, Joseph Paul Cohen, Jiaming Liu, Longchao Liu, Dave Van Veen, Syed Jamal Safdar Gardezi, Hongkun Yu, Magdalini Paschali, Zhihong Chen, et al. Merlin: a computed tomography vision–language foundation model and dataset. *Nature*, 652 (8112):1318–1328, 2026.

Walid Bousselham, Hilde Kuehne, and Cordelia Schmid. VOLD: Reasoning transfer from LLMs to vision-language models via on-policy distillation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26209–26218, 2026.

Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris Coll-Vinent, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, et al. Sam 3: Segment anything with concepts. In *International conference on learning representations*, volume 2026, pp. 138846–138923, 2026.

Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Medical image analysis*, 97:103280, 2024.

Sergios Gatidis, Thomas Küstner, Michael Ingrisch, Tobias Hepp, Marcel Früh, Konstantin Nikolaou, Christian La Fougère, Christina Pfannenberg, Matthias Fabritius, Katharina Jeblick, Balthasar Schachtner, Jakob Dexl, Philipp Wesp, Andreas Mittermeier, Lena Unterrainer, Gabriel Sheikh, Guido Böning, Matthias Brendel, Jens Ricke, Sijing Gu, Thomas Geyer, and Clemens Cyran. Psma-fdg-pet-ct-lesions, May 2024. URL https://doi.org/10.57754/FDAT.6gjsg-zcg93.

Hyunsu Go, Youngung Han, Kyeonghun Kim, Junga Kim, Dohyun Kweon, Jinyong Jun, Sungha Park, Anna Jung, Induk Um, Yului Jeong, et al. Anatomy-privileged distillation with token routing for mri-based prediction of perineural invasion. *arXiv preprint arXiv:2607.11987*, 2026.

Yuxian Gu, Li Dong, Furu Wei, and Minlie Huang. Minillm: Knowledge distillation of large language models. In *The twelfth international conference on learning representations*, 2024.

Ibrahim Ethem Hamamci, Sezgin Er, and Bjoern Menze. Ct2rep: Automated radiology report generation for 3d medical imaging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 476–486. Springer, 2024.

Ibrahim Ethem Hamamci, Sezgin Er, Chenyu Wang, Furkan Almas, Ayse Gulnihan Simsek, Sevval Nil Esirgun, Irem Dogan, Omer Faruk Durugol, Benjamin Hou, Suprosanna Shit, et al. Generalist foundation models from a multimodal dataset for 3d computed tomography. *Nature Biomedical Engineering*, pp. 1–19, 2026.

Ali Hatamizadeh, Vishwesh Nath, Yucheng Tang, Dong Yang, Holger R Roth, and Daguang Xu. Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In *International MICCAI brainlesion workshop*, pp. 272–284. Springer, 2021.

Nicholas Heller, Fabian Isensee, Dasha Trofimova, Resha Tejpaul, Zhongchen Zhao, Huai Chen, Lisheng Wang, Alex Golts, Daniel Khapun, Daniel Shats, Yoel Shoshan, Flora Gilboa-Solomon, Yasmeen George, Xi Yang, Jianpeng Zhang, Jing Zhang, Yong Xia, Mengran Wu, Zhiyang Liu, Ed Walczak, Sean McSweeney, Ranveer Vasdev, Chris Hornung, Rafat Solaiman, Jamee Schoephoerster, Bailey Abernathy, David Wu, Safa Abdulkadir, Ben Byun, Justice Spriggs, Griffin Struyk, Alexandra Austin, Ben Simpson, Michael Hagstrom, Sierra Virnig, John French, Nitin Venkatesh, Sarah Chan, Keenan Moore, Anna Jacobsen, Susan Austin, Mark Austin, Subodh Regmi, Nikolaos Papanikolopoulos, and Christopher Weight. The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct, 2023.

Moritz Roman Hernandez Petzsche, Ezequiel de la Rosa, Roland Wiest, Uta Hanning, Benedikt Wiestler, and Jan S. Kirschke. Isles 2022: A multi-center magnetic resonance imaging stroke lesion segmentation dataset. Zenodo, May 2022. URL `https://doi.org/10.5281/zenodo.7960856`.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015.

Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

Price Jackson, Michael Hofman, James Patrick Buteau, Lachlan McIntosh, and Yu Sun. A repository of annotated psma and fdg pet/ct images for algorithm development in staging of mcrpc for treament with 177lu-psma therapy. Zenodo, November 2025. URL `https://doi.org/10.5281/zenodo.15281784`.

Songtao Jiang, Yuan Wang, Sibo Song, Tianxiang Hu, Chenyi Zhou, Bin Pu, Yan Zhang, Zhibo Yang, Yang Feng, Joey Tianyi Zhou, et al. Hulu-med: A transparent generalist model towards holistic medical vision-language understanding. *arXiv preprint arXiv:2510.08668*, 2025.

Banafsheh Karimian, Alexis Guichemerre, Soufiane Belharbi, Natacha Gillet, Luke McCaffrey, Mohammadhadi Shateri, and Eric Granger. Longitudinal risk prediction in mammography with privileged history distillation. *arXiv preprint arXiv:2603.15814*, 2026.

Yoon Kim and Alexander M Rush. Sequence-level knowledge distillation. In *Proceedings of the 2016 conference on empirical methods in natural language processing*, pp. 1317–1327, 2016.

Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026, 2023.

KiTS23 Challenge Organizers. KiTS23: The 2023 kidney and kidney tumor segmentation challenge. MICCAI 2023 Challenge, 2023. URL `https://kits-challenge.org/kits23/`. Challenge website and dataset description.

Jongwoo Ko, Sungnyun Kim, Tianyi Chen, and Se-Young Yun. Distillm: Towards streamlined distillation for large language models. *arXiv preprint arXiv:2402.03898*, 2024.

Ho Hin Lee, Shunxing Bao, Yuankai Huo, and Bennett A Landman. 3d ux-net: A large kernel volumetric convnet modernizing hierarchical transformer for medical image segmentation. *arXiv preprint arXiv:2209.15076*, 2022.

Wonkyung Lee, Junghyup Lee, Dohyung Kim, and Bumsub Ham. Learning with privileged information for efficient image super-resolution. In *European Conference on Computer Vision*, pp. 465–482. Springer, 2020.

Chunyuan Li, Cliff Wong, Sheng Zhang, Naoto Usuyama, Haotian Liu, Jianwei Yang, Tristan Naumann, Hoifung Poon, and Jianfeng Gao. Llava-med: Training a large language-and-vision assistant for biomedicine in one day. *Advances in neural information processing systems*, 36:28541–28564, 2023.

Menghao Li, Linjie Mu, Yin Wang, Haotian Hu, Yannian Gu, Lujiayi Xue, and Fanyi Wang. Caopd: Confidence-aware on-policy distillation for structured visual prediction. *arXiv preprint arXiv:2609.02401*, 2026a.

Wenxuan Li, Alan Yuille, and Zongwei Zhou. How well do supervised 3d models transfer to medical imaging tasks? In *International Conference on Learning Representations*, volume 2024, pp. 16389–16404, 2024.

Wenxuan Li, Xinze Zhou, Qi Chen, Tianyu Lin, Pedro RAS Bassi, Xiaoxi Chen, Chen Ye, Zheren Zhu, Kai Ding, Heng Li, et al. Pants: The pancreatic tumor segmentation dataset. *Advances in Neural Information Processing Systems*, 38, 2026b.

Yijiang Li, Yijun Liang, Yunjie Tian, Bingyang Wang, Ke Zhang, Zhenfei Yin, Di Fu, Philip Torr, and Nuno Vasconcelos. Self-supervised visual on-policy distillation. *arXiv preprint arXiv:2608.14144*, 2026c.

Alexander Lin, Jeremy Wohlwend, Howard Chen, and Tao Lei. Autoregressive knowledge distillation through imitation learning. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 6121–6133, 2020.

M. W. Linehan, R. Gautam, C. A. Sadow, and S. Levine. The cancer genome atlas kidney chromophobe collection (TCGA-KICH) (version 3) [data set]. The Cancer Imaging Archive, 2016. URL https://doi.org/10.7937/K9/TCIA.2016.YU3RBCZN.

Anglin Liu, Rundong Xue, Xu R Cao, Yifan Shen, Yi Lu, Xiang Li, Qianqian Chen, and Jintai Chen. Medsam3: Delving into segment anything with medical concepts. *arXiv preprint arXiv:2511.19046*, 2025.

Jie Liu, Yixiao Zhang, Jie-Neng Chen, Junfei Xiao, Yongyi Lu, Bennett A Landman, Yixuan Yuan, Alan Yuille, Yucheng Tang, and Zongwei Zhou. Clip-driven universal model for organ segmentation and tumor detection. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 21095–21107. IEEE, 2023.

Ruiqi Liu, Xiaolei Lv, Gengsheng Li, Ximo Zhu, Zhiheng Wang, Zhengbo Zhang, Junkai Chen, Zhiheng Li, Bo Li, Jun Gao, et al. Visual-advantage on-policy distillation for vision-language models. *arXiv preprint arXiv:2605.21924*, 2026.

David Lopez-Paz, Léon Bottou, Bernhard Schölkopf, and Vladimir Vapnik. Unifying distillation and privileged information. *arXiv preprint arXiv:1511.03643*, 2015.

Jiawei Luo, Shixin Huang, Xixi Nie, et al. Primary liver cancer CECT imaging dataset. Science Data Bank, 2025. URL https://cstr.cn/31253.11.sciencedb.12207. CSTR: 31253.11.sciencedb.12207.

Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15:654, 2024a.

Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024b.

Jun Ma, Zongxin Yang, Sumin Kim, Bihui Chen, Mohammed Baharoon, Adibvafa Fallahpour, Reza Asakereh, Hongwei Lyu, and Bo Wang. Medsam2: Segment anything in 3d medical images and videos. *arXiv preprint arXiv:2504.03600*, 2025.

Ahmed W. Moawad, Ayahallah A. Ahmed, Mohab ElMohr, Mohamed Eltaher, Mouhammed Amir Habra, Sarah Fisher, Nancy Perrier, Miao Zhang, David Fuentes, and Khaled Elsayes. Voxel-level segmentation of pathologically-proven adrenocortical carcinoma with Ki-67 expression (Adrenal-ACC-Ki67-Seg). The Cancer Imaging Archive, 2023. URL https://doi.org/10.7937/1FPG-VM46.

Ali Morshid, Khaled M Elsayes, Ahmed M Khalaf, Mohab M Elmohr, Justin Yu, Ahmed O Kaseb, Manal Hassan, Armeen Mahvash, Zhihui Wang, John D Hazle, et al. A machine learning model to predict hepatocellular carcinoma response to transcatheter arterial chemoembolization. *Radiology: Artificial Intelligence*, 1(5):e180021, 2019.

National Cancer Institute Clinical Proteomic Tumor Analysis Consortium. The clinical proteomic tumor analysis consortium clear cell renal cell carcinoma collection (cptac-ccrcc). The Cancer Imaging Archive (TCIA), 2018. URL https://doi.org/10.7937/K9/TCIA.2018.OBLAMN27.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *Transactions on Machine Learning Research Journal*, 2024.

Dmitry Pechyony and Vladimir Vapnik. On the theory of learnining with privileged information. *Advances in neural information processing systems*, 23, 2010.

Dré Peeters, Bogdan Obreja, Noa Antonissen, and Colin Jacobs. The luna25 challenge: Public training and development set - imaging data. Zenodo, February 2025. URL https://doi.org/10.5281/zenodo.14223624.

Fernando Pérez-García, Harshita Sharma, Sam Bond-Taylor, Kenza Bouzid, Valentina Salvatelli, Maximilian Ilse, Shruthi Bannur, Daniel C Castro, Anton Schwaighofer, Matthew P Lungren, et al. Exploring scalable medical image encoders beyond text supervision. *Nature Machine Intelligence*, 7(1):119–130, 2025.

Yunhang Qian, Jiaquan Yu, Jiawei Liu, Meng Wang, Hongwei Bran Li, and Xiaobin Hu. Medopd: Improving medical vision-language models via evidence-aware on-policy distillation. *arXiv preprint arXiv:2607.16303*, 2026.

Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman Rädle, Chloe Rolland, Laura Gustafson, et al. Sam 2: Segment anything in images and videos. In *International Conference on Learning Representations*, volume 2025, pp. 28085–28128, 2025.

Stéphane Ross, Geoffrey Gordon, and Drew Bagnell. A reduction of imitation learning and structured prediction to no-regret online learning. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pp. 627–635. JMLR Workshop and Conference Proceedings, 2011.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, et al. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*, 2024.

Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.

Luc Soler, Alexandre Hostettler, Vincent Agnus, Arnaud Charnoz, Jean-Baptiste Fasquel, Johan Moreau, Anne-Blandine Osswald, Mourad Bouhadjar, and Jacques Marescaux. 3d image reconstruction for comparison of algorithm database. *URL: https://www. ircad. fr/research/datasets/liver-segmentation-3d-ircadb-01*, 13:4, 2010.

Vladimir Vapnik and Rauf Izmailov. Learning using privileged information: similarity control and knowledge transfer. *The Journal of Machine Learning Research*, 16(1):2023–2049, 2015.

Vladimir Vapnik and Akshay Vashist. A new learning paradigm: Learning using privileged information. *Neural networks*, 22(5-6):544–557, 2009.

Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Hui Hui, Yanfeng Wang, and Weidi Xie. Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data. *Nature Communications*, 16(1):7866, 2025.

Linshan Wu, Jiaxin Zhuang, and Hao Chen. Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22873–22882. IEEE, 2024.

Linshan Wu, Jia-Xin Zhuang, and Hao Chen. Glance and focus reinforcement for pan-cancer screening. In *International Conference on Learning Representations*, volume 2026, pp. 1998–2021, 2026.

Yutong Xie, Jianpeng Zhang, Yong Xia, and Qi Wu. Unimiss+: Universal medical self-supervised learning from cross-dimensional unpaired data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10021–10035, 2024.

Weiwen Xu, Hou Pong Chan, Long Li, Mahani Aljunied, Ruifeng Yuan, Jianyu Wang, Chenghao Xiao, Guizhen Chen, Chaoqun Liu, Zhaodonghui Li, et al. Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning. *arXiv preprint arXiv:2506.07044*, 2025.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Hangjie Yuan, Yichen Qian, Zhiwei Tang, Xianzhe Xu, Lirong Wu, Sicheng Yang, Jinwang Wang, Pengju Wang, Zhitao Zeng, Yizeng Han, et al. Clinfusion: A vision-centric multimodal llm system for holistic medical understanding. *arXiv preprint arXiv:2607.24743*, 2026a.

Qianhao Yuan, Jie Lou, Xing Yu, Hongyu Lin, Le Sun, Xianpei Han, and Yaojie Lu. Vision-opd: Learning to see fine details for multimodal llms via on-policy self-distillation. *arXiv preprint arXiv:2605.18740*, 2026b.

Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, et al. Biomedclip: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. *arXiv preprint arXiv:2303.00915*, 2023.

Peisen Zhao, Lingxi Xie, Jiajie Wang, Ya Zhang, and Qi Tian. Progressive privileged knowledge distillation for online action detection. *Pattern Recognition*, 129:108741, 2022.

Tong Zhao, Yuyang Hu, Yutao Zhu, Reed Li, Haijin Liang, Haibo Shi, Yu Lu, and Zhicheng Dou. Opod: On-policy omni distillation. *arXiv preprint arXiv:2607.20918*, 2026.

Jiayuan Zhu, Abdullah Hamdi, Yunli Qi, Yueming Jin, and Junde Wu. Medical sam 2: Segment medical images as video via segment anything model 2. *arXiv preprint arXiv:2408.00874*, 2024.

A BENCHMARK DETAILS

A.1 BENCHMARK SCOPE AND TASK SETTINGS

We construct MedBridge-Bench with 20,689 scored instances from 18,595 unique studies to evaluate the distilled student alongside task-specific models and generalist medical foundation models. It comprises five sub-benchmarks, *MedBridge-CT.TumorExt*, *MedBridge-CT.TumorInt*, *MedBridge-CT.Nodule*, *MedBridge-MR*, and *MedBridge-PET/CT* (abbreviated as *M.B.* in tables), assembled from 20 public datasets and three private clinical cohorts spanning CT, MRI (DWI), and PET/CT.

- **Studies and scored instances.** A *study* denotes one 3D image series and is used interchangeably with *case* and, in the dataset descriptions, *volume*; one patient may contribute multiple studies. A *scored instance* normally corresponds to one study. The exception is *MedBridge-CT.Nodule*, where 4,069 thin-slice chest CT studies yield 6,163 nodule-level instances because each study may contain multiple annotated nodules. Some sources are listed twice in Tab. 6 but counted only once in the benchmark totals. The table therefore sums to 23,713 instances, 3,024 more than the de-duplicated total of 20,689; the same difference applies to study counts.
- **Task design.** Two output levels are used. Case-level lesion detection matches clinical reading and is reported wherever a cohort supplies case- or lesion-level labels; voxel-level segmentation is reported wherever masks exist, as it distinguishes accurate lesion localization from responses to merely plausible regions and naturally measures knowledge transfer from dense visual teachers. Which of the two levels applies is stated in the Task paragraph of each sub-benchmark below. *MedBridge-CT.Nodule* further evaluates nodule-level benign-versus-malignant classification, requiring fine-grained morphological reasoning and allowing multiple scored instances per study.
- **Sub-benchmark roles.** *MedBridge-CT.TumorExt* combines in-distribution training and evaluation with both IID and organ-specific OOD cohorts to measure transfer, whereas *MedBridge-CT.TumorInt* evaluates multiple tumor classes within a unified pan-cancer collection. Both include lesion-free controls for clinically meaningful detection evaluation. *MedBridge-MR* evaluates ischemic-stroke lesion analysis in DWI images, while *MedBridge-PET/CT* provides the only cross-modal setting and comprises an mCRPC segmentation task (*MedBridge-PET/CT-Seg*) and a whole-body classification task over four tumor types and lesion-free controls (*MedBridge-PET/CT-Cls*).
- **Evaluation.** For multi-class datasets, we report class-wise scores together with macro averages to characterize both category-specific performance and overall performance without weighting classes by their sample frequencies. Detection is evaluated using the threshold-free AUC and the threshold-dependent sensitivity, specificity, F1, and balanced accuracy (BAcc). Unless otherwise noted, threshold-dependent metrics are computed at each model's Youden-index operating point; a fixed threshold of 0.5 is used where specified. We additionally report $\mathrm{Se}_{@95}$, the sensitivity attained at specificity $\geq 95\%$, and $\mathrm{Sp}_{@95}$, the specificity attained at sensitivity $\geq 95\%$, reflecting clinically relevant screening and rule-out operating points. Segmentation is evaluated using Dice for regional overlap and, where masks are available, normalized surface Dice at a 2 mm tolerance and the 95th-percentile Hausdorff distance (HD95; mm; lower is better). The latter metrics characterize boundary agreement and worst-case boundary errors beyond volumetric overlap. Dice is averaged over lesion-positive cases only.
- **Benchmark composition.** Tab. 6 summarizes the instance count, modality, anatomical region, disease, train/test split, and data source for each sub-benchmark.

A.2 MEDBRIDGE-CT.TUMOREXT

Except where noted, every source below is used as released: we re-derive labels and case-level targets but do not modify the imaging.

Task. Case-level tumor detection in liver, kidney and pancreas, each lesion class scored one-vs-rest against lesion-free controls, together with voxel-level segmentation wherever masks exist. Evaluation has two arms: an in-distribution (IID) arm on which models are trained and internally evaluated,

Table 6: Composition of MedBridge-Bench: instance count, modality, anatomical region, disease, split, and source for each sub-benchmark.

Sub-benchmark / dataset	Modality	Region – disease / lesion class	Instances (train/ test)	Source
M.B.-CT.TumorExt (IID)	CT	Liver, kidney, pancreas – tumor	9,262 (8,336 / 926)	AbdomenAtlas3.0 (Bassi et al., 2025)
M.B.-CT.TumorExt (OOD)	CT	Liver, pancreas, kidney – tumor	1,783 (– / 1,783)	WAW-TACE (Bartnik et al., 2024), 3D-IRCADb-01 (Soler et al., 2010), PLC-CECT (Luo et al., 2025), PanTS (Li et al., 2026b), TCGA-KIRC (Akin et al., 2016), TCGA-KICH (Linehan et al., 2016), CPTAC-CCRCC (National Cancer Institute Clinical Proteomic Tumor Analysis Consortium, 2018), Private (single center)
M.B.-CT.TumorInt	CT	Lung, liver, kidney, pancreas, colon, adrenal – tumor	4,541 (3,827 / 714)	Pancancer (Wu et al., 2026), NSCLC-Radiomics (Aerts et al., 2014), NSCLC-Radiogenomics (Bakr et al., 2017), MSD (Medical Segmentation Decathlon) (Antonelli et al., 2022), WAW-TACE (Bartnik et al., 2024), HCC (Morshid et al., 2019), PANORAMA (Alves et al., 2024), KiTS23 (Heller et al., 2023), Adrenal-Ki67-Seg (Moawad et al., 2023), AbdomenAtlas (Bassi et al., 2025)
M.B.-CT.Nodule	CT	Lung – pulmonary nodule, benign / malignant	6,163 (4,892 / 1,271)	LUNA25 (Peeters et al., 2025)
M.B.-MR	MRI (DWI)	Brain – acute ischemic stroke	250 (200 / 50)	ISLES'22 (Hernandez Petzsche et al., 2022)
M.B.-PET/CT-Seg	CT + PSMA	Prostate / pelvis – mCRPC	100 (80 / 20)	DEEP-PSMA (Jackson et al., 2025)
M.B.-PET/CT-Cls	CT + FDG / PSMA	Whole-body – melanoma, lung ca., lymphoma, prostate ca.; lesion-free	1,614 (1,291 / 323)	PSMA-FDG-PET-CT-Lesions (Gatidis et al., 2024)

and an out-of-distribution (OOD) arm of three organ-specific external cohorts that is held out entirely, so this sub-benchmark carries the transfer measurements.

Split. The IID arm contains 9,262 scored instances, 8,336 for training and 926 for evaluation, following the split released with AbdomenAtlas3.0 (Tab. 6). The OOD arm contains 1,783 evaluation-only cases that overlap neither IID partition: 650 liver, 750 pancreas and 383 kidney, with per-cohort counts and contrast-phase composition in Tab. 7. For PanTS we removed every case overlapping the IID partition, leaving a leakage-free subset of 700. Case-level decision thresholds were determined once for each model using the Youden index on the IID cohort and were subsequently frozen for all OOD evaluation. Detection and segmentation metrics are reported for the liver and pancreas cohorts, whereas the kidney cohort is evaluated at the case level using BAcc, AUC and F1; kidney-tumor Dice is not reported because the selected external kidney cohorts do not provide compatible voxel-level tumor masks. Dice is averaged over lesion-positive cases only.

Data sources. The IID arm draws on a single source. **AbdomenAtlas3.0** (Bassi et al., 2025) is the tumor-centric abdominal CT corpus released with RadGPT, covering liver, kidney and pancreatic tumors. It pools many public abdominal cohorts from multiple institutions and is therefore heterogeneous in scanner, reconstruction kernel, slice thickness and contrast phase; tumor masks are semi-automatic and revised by board-certified radiologists, and each study carries a radiology report. The OOD arm draws on seven public cohorts and one private set across the three organ arms. *Liver:* **WAW-TACE** (Bartnik et al., 2024) is a single-center hepatocellular-carcinoma cohort from the Medical University of Warsaw treated with transarterial chemoembolization, providing four contrast phases (non-contrast, arterial, portal-venous, delayed) with hand-drawn tumor segmentations, organ masks, precomputed radiomics and clinical data. **3D-IRCADb-01** (Soler et al., 2010) is an early public 3D liver CT benchmark from IRCAD Strasbourg with manual expert ground truth for liver, lesions, vasculature and biliary tree over arterial and portal-venous phases, covering HCC, cholangiocarcinoma, metastases and cysts. **PLC-CECT** (Luo et al., 2025) is a single-center four-phase resource (plain, arterial, venous, delayed) built for subtype-level differential diagnosis of HCC versus ICC versus cHCC-CCA, with liver and lesion masks. *Pancreas:* **PanTS** (Li et al., 2026b) is a large multi-institutional pancreatic tumor segmentation set with expert-validated voxel-wise labels that separates PDAC, neuroendocrine tumors, cystic neoplasms and non-neoplastic cysts, shipped with nnU-Net baselines. *Kidney:* **TCGA-KIRC** (Akin et al., 2016) is the TCGA imaging

Table 7: Data sources and case-level composition of the external cohorts (Tab. 2). Values denote case counts. Pooled rows sum all listed cohorts (public and private). Phase abbreviations: A = arterial, V = venous/portal, D = delayed, NC = non-contrast, CE = contrast-enhanced. †Phase partially recorded in source DICOMs.

Organ	Cohort	Modality	Phase	Pathology	Pos.	Neg.	Total
Liver	WAW-TACE (Bartnik et al., 2024)	CT	A 118 / V 69 / D 29 / NC 3	HCC	219	0	219
	3D-IRCADb-01 (Soler et al., 2010)	CT	CE†	HCC, Liver metastases	15	5	20
	PLC-CECT (Luo et al., 2025)	CT	CE	HCC, ICC, cHCC-CCA	278	83	361
	Private	CT	NC	–	0	50	50
	Liver pooled	CT	Mixed		**512**	**138**	**650**
Pancreas	PanTS (leakage-free) (Li et al., 2026b)	CT	V 265 / NC 289 / A 139 / D 7	PDAC, pNET, cystic neoplasm, cystic non-neoplastic lesion	186	514	700
	Private	CT	NC	–	0	50	50
	Pancreas pooled	CT	Mixed		**186**	**564**	**750**
Kidney	TCGA-KIRC (Akin et al., 2016)	CT	Mixed†	Clear Cell Carcinoma	235	0	235
	TCGA-KICH (Linehan et al., 2016)	CT	Mixed†	Chromophobe	12	0	12
	CPTAC-CCRCC (National Cancer Institute Clinical Proteomic Tumor Analysis Consortium, 2018)	CT	Mixed†	Clear Cell Carcinoma	61	0	61
	Private	CT	NC	–	0	75	75
	Kidney pooled	CT	Mixed		**308**	**75**	**383**

companion for clear-cell renal cell carcinoma, joining CT and MR to histopathology, genomic and clinical records by shared case identifier. **TCGA-KICH** (Linehan et al., 2016) is the corresponding companion for chromophobe renal cell carcinoma, distinguished by the same radiogenomic linkage rather than by dense lesion annotation. **CPTAC-CCRCC** (National Cancer Institute Clinical Proteomic Tumor Analysis Consortium, 2018) provides clear-cell RCC imaging from NCI's Clinical Proteomic Tumor Analysis Consortium, linkable to large-scale proteomic measurements, with lesion-level contours supplied in a companion release. All three arms additionally draw lesion-negative controls from the private cohort described under Ethical considerations below.

Ethical considerations. The public cohorts were accessed and used in accordance with their respective data-use agreements and licenses. To provide lesion-negative controls for external case-level evaluation, we additionally collected 175 retrospective cases from partner hospitals: 50 for liver, 50 for pancreas, and 75 for kidney. Collection and use of these cases were approved by the relevant institutional review board (IRB). All private scans were de-identified before analysis and were used exclusively for external evaluation; they were not used for model training, checkpoint selection, hyperparameter tuning, or decision-threshold calibration.

A.3 MEDBRIDGE-CT.TUMORINT

Task. Lesion-level classification over six tumor classes spanning lung, liver, kidney, pancreas, colon and adrenal within one unified pan-cancer collection, reported per class and as macro averages, with lesion-free controls so that detection stays clinically meaningful. Evaluation is internal only.

Split. 4,541 scored instances: 3,827 for training and 714 for evaluation (Tab. 6).

Data sources. Ten sources are pooled. The **Pan-cancer Dataset** (Wu et al., 2026) is an aggregated chest-and-abdomen CT screening corpus with voxel-level lesion masks over nine categories and radiology-report-verified lesion-free controls. **NSCLC-Radiomics** (Aerts et al., 2014) is the imaging backbone of the study that decoded tumor phenotype noninvasively: pretreatment radiotherapy-planning CT with gross-tumor-volume delineations by radiation oncologists and clinical outcome data, now the canonical public radiomics benchmark. **NSCLC-Radiogenomics** (Bakr et al., 2017) extends the same NSCLC cases with PET/CT and paired multi-omic annotations (EGFR, KRAS, ALK, RNA-seq) alongside voxel-level tumor segmentations. **MSD** (Antonelli et al., 2022), the Medical Segmentation Decathlon, comprises ten multi-institutional organ-lesion segmentation tasks in CT and MRI, three of them withheld as unseen mysteries. **HCC** (Morshid et al., 2019) originates from a single-center retrospective cohort at MD Anderson Cancer Center assembled to predict HCC response to TACE, and provides pre-therapeutic contrast-enhanced CT with liver and tumor segmentations plus BCLC staging and time-to-progression outcome data. **PANORAMA** (Alves et al., 2024) is an international PDAC detection challenge on standard-of-care contrast-enhanced CT whose reference standard is histopathology with at least three years of follow-up. **KiTS23** (KiTS23 Challenge Organizers, 2023) is the third iteration of the Kidney and Kidney Tumor Segmentation challenge, with expert-validated masks for kidneys, renal tumors and cysts on contrast-enhanced preoperative CT and a held-out test set. **Adrenal-ACC-Ki67-Seg** (Moawad et al., 2023) pairs voxel-level adrenocortical-carcinoma segmentations with the histopathological Ki-67 proliferation index

for imaging-based grade and prognosis prediction. **WAW-TACE** (Bartnik et al., 2024) and **AbdomenAtlas3.0** (Bassi et al., 2025) are also included, both described under *MedBridge-CT.TumorExt*, which is why Tab. 6 lists them twice while the benchmark totals count them once.

A.4 MEDBRIDGE-CT.NODULE

Task. Nodule-level benign-versus-malignant classification on crops centered at each annotated nodule, which demands fine-grained morphological reasoning and is the only setting in MedBridge-Bench where a single study yields several scored instances.

Split. 6,163 nodule-level instances drawn from 4,069 thin-slice chest CT studies: 4,892 for training and 1,271 for evaluation (Tab. 6). The partition is made at patient level and stratified by malignancy, so that no patient appears on both sides.

Data source. **LUNA25** (Peeters et al., 2025) is the successor to LUNA16, built on retrospective multi-center low-dose screening chest CT from the National Lung Screening Trial. Its target is malignancy risk per lesion rather than detection or segmentation, and it supports direct comparison against radiologist performance.

A.5 MEDBRIDGE-MR

Task. Ischemic-stroke lesion analysis in diffusion-weighted MRI, evaluated at voxel level with Dice, normalized surface Dice and HD95.

Split. 250 scored instances: 200 for training and 50 for evaluation (Tab. 6). The partition is fixed at patient level and stratified by lesion volume.

Data source. **ISLES'22** (Hernandez Petzsche et al., 2022) targets acute-to-subacute ischemic stroke lesion segmentation and is multi-center and multi-vendor (1.5 T and 3 T Siemens/Philips), providing FLAIR, DWI, ADC and expert-corrected lesion masks whose size, number and location vary widely across cases. Its public test labels are unavailable, which is why we fix our own patient-level split. Only DWI is used, because the sequences are not inter-registered.

A.6 MEDBRIDGE-PET/CT

Task. The only cross-modal setting in MedBridge-Bench, with CT and PET read jointly, and the only one that splits into two tasks. *MedBridge-PET/CT-Seg* segments mCRPC in the prostate and pelvis and is scored with Dice, normalized surface Dice and HD95. *MedBridge-PET/CT-Cls* classifies whole-body studies over four tumor types (melanoma, lung cancer, lymphoma, prostate cancer) against lesion-free controls and is scored with BAcc, AUC and F1, with Dice additionally reported from the available manual segmentations.

Split. *-Seg*: 100 scored instances, 80 for training and 20 for evaluation. *-Cls*: 1,614 scored instances, 1,291 for training and 323 for evaluation (Tab. 6). Both partitions are patient-disjoint: every fifth patient is assigned to the *-Seg* validation side, and the *-Cls* split groups 1,023 training against 255 evaluation patients with zero overlap, 247 of them contributing more than one study.

Data sources. **DEEP-PSMA** (Jackson et al., 2025) supplies *-Seg*: dual-tracer (PSMA + FDG) mCRPC PET/CT from Peter MacCallum Cancer Centre with independent expert contours drawn separately for PSMA-avid and FDG-avid disease, enabling a direct tracer-avidity comparison. We keep the PSMA studies only, giving 100 CT + PET pairs. **PSMA-FDG-PET-CT-Lesions** (Gatidis et al., 2024) supplies *-Cls*: a Tübingen whole-body PET/CT repository with two non-paired cohorts, an FDG arm covering melanoma, lymphoma and lung cancer and a PSMA arm covering prostate cancer, with manually drawn tumor segmentations and explicit lesion-free controls. The case-level label is derived as "any tumor voxel present".

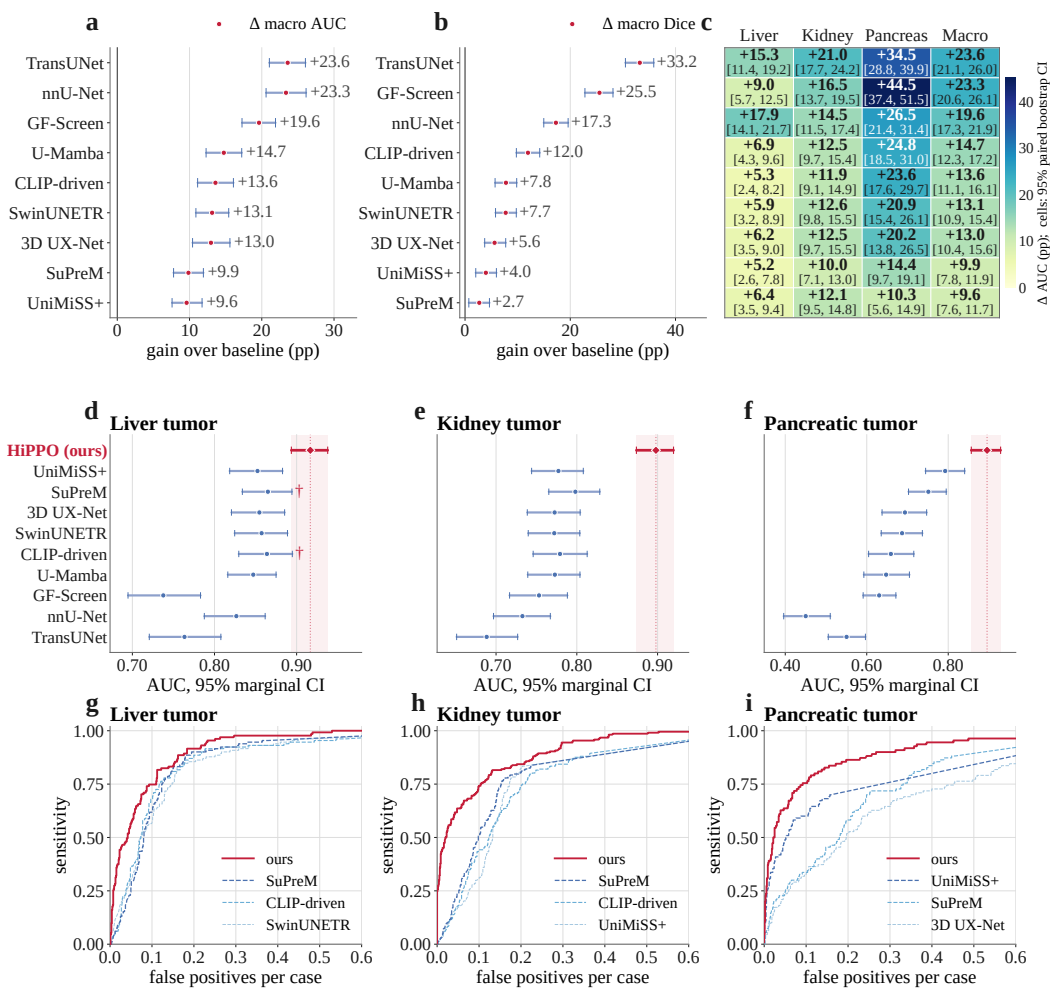

Figure 5: Statistical analysis on the internal test set of *MedBridge-CT.TumorExt*. **a.** Macro-average ΔAUC over three organs (95% paired bootstrap CI, N=2,000). **b.** Macro-average ΔDice with 95% paired bootstrap CI (N=2,000). **c.** Per-organ and macro ΔAUC with 95% bootstrap CIs. **d–f.** Per-organ lesion-level AUC per model (95% marginal bootstrap CI, N=2,000). Shaded band and dotted line mark HiPPO; † denotes marginal-CI overlap with HiPPO (a weaker criterion than the paired tests in **a–c**). **g–i.** Per-organ case-level FROC curves. All 27 paired DeLong comparisons and nine ΔDice tests remain significant after Holm–Bonferroni correction.

B Experiments Details

B.1 Additional Results and Analysis

Statistical analysis. Fig. 5 answers a question a table of point estimates cannot: whether the ordering survives resampling of the cases. Every interval in panels **a–f** is a case-level percentile bootstrap with $N = 2{,}000$ resamples at a fixed seed; the paired intervals of panels **a–c** resample the same cases for both models, so an interval excluding zero states that HiPPO wins on the same cases rather than merely on average. Panels **d–i** then separate ranking from usability: the ROC panels show that the ordering is not produced by a single region of the curve, and the FROC panels price each operating point in false positives per case, which is the quantity a screening program actually pays. The correction is applied over all paired comparisons jointly, so the family-wise statement covers every organ–metric pair rather than each comparison in isolation; all 27 paired DeLong comparisons and the nine ΔDice tests remain significant after Holm–Bonferroni correction.

Figs. 5 and 6 report paired per-case gains of HiPPO over every baseline, with 95% bootstrap intervals and Holm–Bonferroni correction within each comparison family. Across 82 paired compar-

isons, every point estimate favors HiPPO, and 67 remain significant after correction. The evidence is strongest on *MedBridge-CT.TumorExt*, where all 36 comparisons are significant, and the improvements on the larger cohorts are often substantial rather than merely statistically detectable. The paired analysis thus shows a consistent direction of improvement across the benchmark.

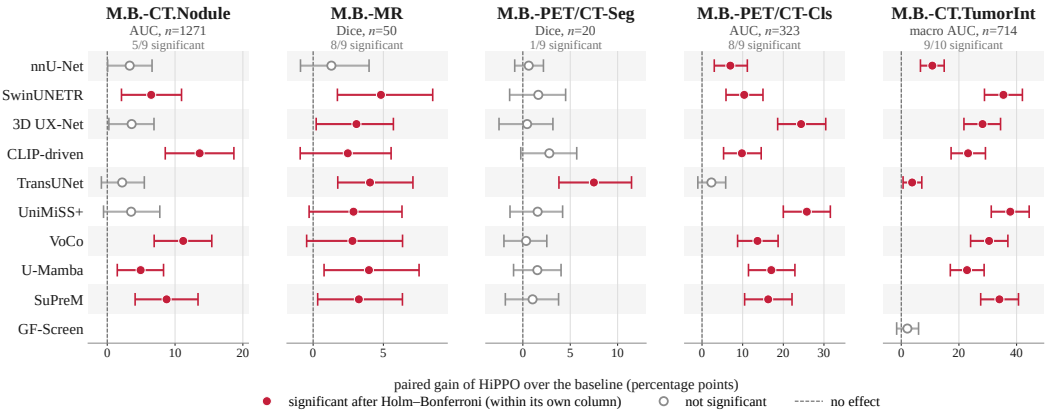

Figure 6: Paired gains of HiPPO over every baseline on the five external sub-benchmarks, drawn as one matrix: rows are the baselines in table order, columns are the sub-benchmarks (the AUC and Dice columns of Tab. 4 and the macro AUC column of Tab. 3), each with its own horizontal scale. Whiskers are 95% paired bootstrap CIs (N=10,000); filled markers survive Holm–Bonferroni correction within their own column (paired DeLong for the AUC columns, Wilcoxon signed-rank for the Dice columns) and hollow markers do not. Cohort size and the per-column significance count are given in the column headers; GF-Screen competes only on *MedBridge-CT.TumorInt*.

Visualization. HiPPO attains strong lesion segmentation performance on MedBridge-CT.TumorExt, and Fig. 7 provides several illustrative examples that juxtapose our predictions with those of the nine baselines, reporting both whole-volume and slice-level Dice for each method.

External performance across lesion sizes and operating regimes. Tab. 8 and Tab. 9 show that HiPPO maintains a stronger sensitivity–specificity balance across lesion sizes and operating regimes. At the transferred thresholds, it achieves 83.98% sensitivity with 85.51% specificity on liver tumors and 68.28% sensitivity with 85.64% specificity on pancreatic tumors, while several baselines improve one metric only by sacrificing the other. The advantage remains at clinically demanding ROC endpoints: at 95% specificity, HiPPO reaches 75.98% liver sensitivity, compared with 32.62% for the strongest baseline; at 95% sensitivity, it retains 57.97% liver, 12.00% kidney and 32.98% pancreatic specificity, whereas every baseline falls to zero on liver and kidney and the best baseline reaches only 8.51% on pancreas. Its strong small-lesion performance further indicates that the gain is not restricted to large, conspicuous lesions or a favorable operating threshold.

Table 8: Case-level sensitivity stratified by ground-truth lesion size, the organ-wide sensitivity, and specificity, on the OOD test set of MedBridge-CT.TumorExt.{organ}.

Model	Liver Tumor (%)				Kidney Tumor (%)		Pancreas Tumor (%)			
	$Se_{\leq 2\,cm}$	$Se_{>2\,cm}$	Se_{all}	Spec	Se_{all}	Spec	$Se_{\leq 2\,cm}$	$Se_{>2\,cm}$	Se_{all}	Spec
nnU-Net (Isensee et al., 2021)	38.89	51.42	50.98	84.78	83.44	57.33	**95.16**	**95.16**	**95.16**	8.51
SwinUNETR (Hatamizadeh et al., 2021)	22.22	47.37	46.48	91.30	75.00	81.33	41.94	61.29	54.84	72.34
3D UX-Net (Lee et al., 2022)	44.44	61.54	60.94	87.68	75.65	89.33	35.48	58.87	51.08	81.74
CLIP-driven (Liu et al., 2023)	22.22	49.60	48.63	73.91	**85.39**	29.33	56.45	55.65	55.91	48.23
TransUNet (Chen et al., 2024)	5.56	38.26	37.11	93.48	64.29	92.00	8.06	20.97	16.67	93.09
UniMiSS+ (Xie et al., 2024)	38.89	57.29	56.64	89.13	82.47	94.67	16.13	68.55	51.08	88.12
U-Mamba (Ma et al., 2024b)	22.22	36.44	35.94	94.20	83.12	92.00	4.84	36.29	25.81	96.63
SuPreM (Li et al., 2024)	44.44	57.69	57.23	88.41	65.91	94.67	37.10	62.90	54.30	77.84
GF-Screen (Wu et al., 2026)	5.56	32.39	31.45	**98.55**	71.10	**100.00**	4.84	26.61	19.35	**99.29**
HiPPO (ours)	**72.22**	**84.41**	**83.98**	85.51	84.42	93.33	54.84	75.00	68.28	85.64

Additional ablation studies. Tab. 10 isolates the remaining design choices under the same training schedule and supervision budget; every variant differs from HiPPO (full) in exactly one mecha-

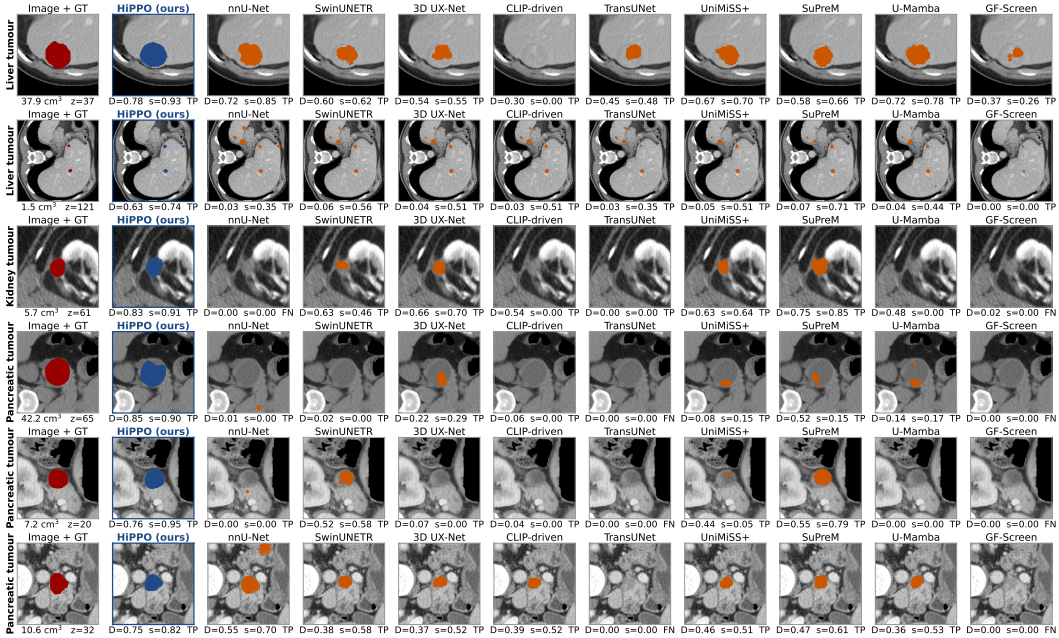

Figure 7: Qualitative lesion segmentation on the internal test set of MedBridge-CT.TumorExt. Each row is one case with a ground-truth (GT) lesion on the axial slice with the largest GT lesion area; The first column overlays the GT as a solid dark-red mask on the CT; each model column shows only that model's own prediction, solid dark blue for HiPPO and solid dark orange for a baseline; D is the whole-volume 3D Dice, s the Dice restricted to the displayed slice.

Table 9: Case-level operating characteristics read from the ROC curves on the OOD test cohorts of MedBridge-CT.TumorExt.{organ}: sensitivity at specificity $\geq 95\%$ and specificity at sensitivity $\geq 95\%$. Sensitivity is stratified by ground-truth lesion size for the liver and pancreas cohorts, where voxel-level masks are available; only overall sensitivity is reported for the kidney cohort.

Model	Liver Tumor (%)				Kidney Tumor (%)		Pancreas Tumor (%)			
	$\text{Se}_{@95}$			$\text{Sp}_{@95}$	$\text{Se}_{@95}$	$\text{Sp}_{@95}$	$\text{Se}_{@95}$			$\text{Sp}_{@95}$
	≤ 2 cm	> 2 cm	**All**				≤ 2 cm	> 2 cm	**All**	
nnU-Net (Isensee et al., 2021)	16.67	27.33	26.95	0.00	20.13	0.00	1.61	6.45	4.84	<u>8.51</u>
SwinUNETR (Hatamizadeh et al., 2021)	16.67	21.86	21.68	0.00	33.44	0.00	<u>14.52</u>	22.58	19.89	0.00
3D UX-Net (Lee et al., 2022)	27.78	27.94	27.93	0.00	27.27	0.00	9.68	25.81	20.43	0.00
CLIP-driven (Liu et al., 2023)	5.56	12.15	11.91	0.00	35.71	0.00	0.00†	0.00†	0.00†	0.00
TransUNet (Chen et al., 2024)	5.56	25.10	24.41	0.00	58.12	0.00	6.45	18.55	14.52	0.00
UniMiSS+ (Xie et al., 2024)	22.22	17.41	17.58	0.00	57.14	0.00	<u>14.52</u>	**59.68**	**44.62**	0.00
U-Mamba (Ma et al., 2024b)	22.22	27.53	27.34	0.00	61.04	0.00	4.84	47.58	<u>33.33</u>	0.00
SuPreM (Li et al., 2024)	<u>33.33</u>	32.59	<u>32.62</u>	0.00	64.29	0.00	4.84	20.16	15.05	0.00
GF-Screen (Wu et al., 2026)	5.56	<u>33.00</u>	32.03	0.00	**72.40**	0.00	4.84	27.42	19.89	0.00
HiPPO (ours)	**50.00**	**76.92**	**75.98**	**57.97**	<u>70.13</u>	**12.00**	**25.81**	<u>54.03</u>	**44.62**	**32.98**

nism, so each row measures how much that component contributes. *(a) Teacher-reward aggregation.* Mean aggregation outperforms soft-max aggregation at both temperatures, by 1.56, 1.82, and 3.46 points of AUC, BAcc, and F1 over soft-max at $\tau = 0.1$ (the gaps to $\tau = 1.0$ are $1.73/1.62/4.10$), indicating that retaining feedback from all applicable teachers is more effective than concentrating the reward on the strongest response. *(b) Spatial alignment granularity.* The sample-level row is the conventional alternative used in prior feature-distillation work, not our mechanism. It mean-pools the student bottleneck and each teacher's aligned target over all positions $p \in \Omega$ of a crop into one vector each and matches the two pooled vectors with a single cosine distance, so the teacher–student discrepancy collapses to one scalar per crop and every position inside that crop receives identical supervision; the three criteria of Eq. 10 are still formed per position and then averaged over Ω, which keeps the weighting identical in composition and makes granularity the only variable. Our position-wise matching instead resolves the discrepancy at every bottleneck position p, and improves AUC, BAcc, and F1 by 1.79, 1.88, and 1.24 points over this conventional alternative, the gain is attributable to our spatially fine-grained alignment itself. *(c) Dynamic spatial weighting.*

The three criteria entering the position-wise maximum are complementary, and each is necessary: replacing the realized weight by a uniform one costs $1.44/1.89/4.01$ points of AUC/BAcc/F1, and dropping any single criterion degrades all three metrics, the boundary term G_p by $1.43/1.66/2.85$, the difficulty term $D_{k,p}$ by $1.71/1.51/2.07$, and the segmentation-error term η_p by $3.20/3.20/4.74$, the largest decline in the table. That η_p matters most although it is nonzero on only a thin sliver of positions is expected: it upweights precisely the few positions where the current segmentation contradicts the teachers, and those positions decide the boundary and small-lesion pixels. No single criterion suffices and none is dispensable. Overall, the full configuration achieves the best result on every reported macro metric, supporting the joint use of mean teacher aggregation, position-wise distillation, and dynamic spatial weighting.

Table 10: Additional mechanism ablations on the internal test set of *MedBridge-CT.TumorExt*. Results are macro averages over the three lesion classes. Ablations reported in the main text are omitted. Best results are shown in bold.

Variant	AUC	BAcc	F1
(a) Teacher-reward aggregation			
Soft-max reward, $\tau = 1.0$	88.60	81.92	57.19
Soft-max reward, $\tau = 0.1$	88.77	81.72	57.83
(b) Spatial alignment granularity			
Sample-level distillation (conventional)	88.54	81.66	60.05
(c) Dynamic spatial weighting			
Uniform position weight	88.89	81.65	57.28
w/o boundary term (G_p)	88.90	81.88	58.44
w/o difficulty term ($D_{k,p}$)	88.62	82.03	59.22
w/o error term (η_p)	87.13	80.34	56.55
HiPPO (full)	**90.33**	**83.54**	**61.29**

B.2 Summary: Feature-Space Changes from Distillation

Probe construction. We evaluate the six encoder stages *s0–s5*, ordered from fine to coarse. For each case and stage, slice features are aggregated into a 3D grid whose cells are labeled using ground-truth lesion coverage. The same cells are sampled for every model, with negatives drawn either from all background or only from healthy parenchyma of the target organ. Features are projected to $d = 128$, and an L_2-regularized logistic probe is evaluated using five-fold case-level cross-validation. We report mean per-case AUC with 95% bootstrap intervals and use paired bootstrap tests with Holm–Bonferroni correction over the twelve stage-by-negative-mode comparisons. The usable cohort decreases at coarse stages because fewer cells satisfy the positive and negative sampling requirements.

Distillation acts on the deeper stages. This depth-resolved probe is a separate experiment from the facet probe of Fig. 4: it uses its own cell lattice and its own $d = 128$ readout, so its absolute values are not interchangeable with the numbers reported there, and comparisons are meaningful only between models within the same stage. The two trained students are indistinguishable at the finest stage, and the undistilled student is marginally ahead at the pre-registered *s2* endpoint. This is consistent with where the method injects its supervision: the alignment targets are read at bottleneck resolution, so the transferred signal is concentrated in the coarse stages rather than in the shallow, largely textural ones. The ordering reverses from *s3* onward and the gap widens at *s4* and *s5*, where the same-organ contrast shows the largest improvement. After Holm–Bonferroni correction, significance is attained at both ends of the depth axis, comprising shallow-stage reductions and coarse-stage gains.

Relation to the frozen teachers. On the common lattice in Fig. 8**c**, every model is scored by one shared readout, and the distilled student attains the highest coarse-stage probe AUC of the three student arms, falling inside the range spanned by the frozen teachers while the random-initialized control stays below it. This supports the view that distillation makes deeper student features more linearly aligned with the teachers' lesion-related representations.

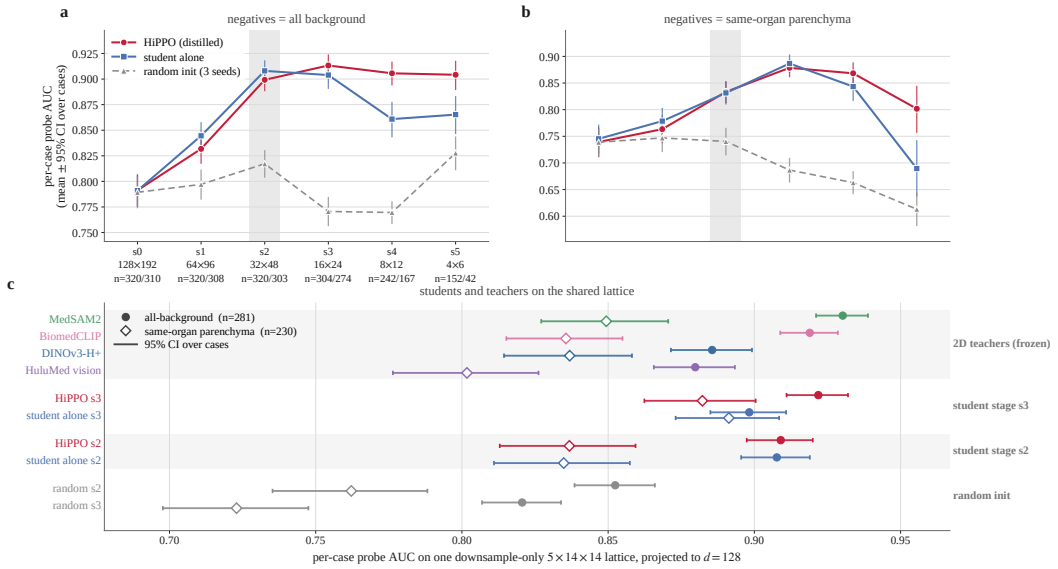

Figure 8: Linear-probe lesion separability across the 3D encoder on the internal test set of *MedBridge-CT.TumorExt*. **a,b.** Per-case AUC at stages s0–s5 for HiPPO, the undistilled student, and a random-initialized control, using all-background (**a**) or same-organ (**b**) negatives. Markers and whiskers denote means and 95% bootstrap CIs; the gray band marks the pre-registered endpoint s2. **c.** Student stages and frozen 2D teachers evaluated on a common lattice with the same $d = 128$ linear readout. This probe uses a different lattice and readout from Fig. 4, so absolute values should be compared within stages; probe AUC measures linearly readable lesion evidence, not segmentation performance.

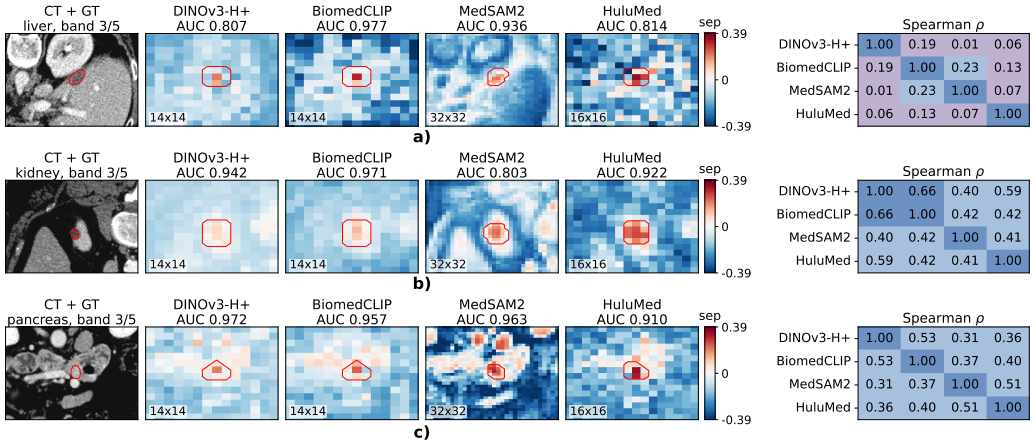

Figure 9: Frozen teachers provide strong but spatially distinct lesion evidence. Rows a–c show three sampled training cases, each with a lesion-centered CT slice, the ground-truth contour in red, and the corresponding teacher-specific patch-separability maps on a shared color scale. Map titles report per-case AUC, and the right column reports the cross-teacher Spearman correlation. Liver is selected as a low-agreement case with consistently high teacher AUCs; kidney and pancreas are selected near their median separability. Cohort-level analyses are provided in Fig. 4.

B.3 Summary: Separability and Multi-Expert Complementarity

The separability analyses first examine whether the frozen supervision contains usable lesion-related information and whether that information is redundant across teachers. On a randomly sampled subset of the training set, we construct each teacher's patch-level separability map as $\text{sep} = \cos(f, c_{\text{fg}}) - \cos(f, c_{\text{bg}})$, using the case-specific lesion and background centroids; three such cases are displayed in Fig. 9. The teachers achieve strong lesion–background separation, but their

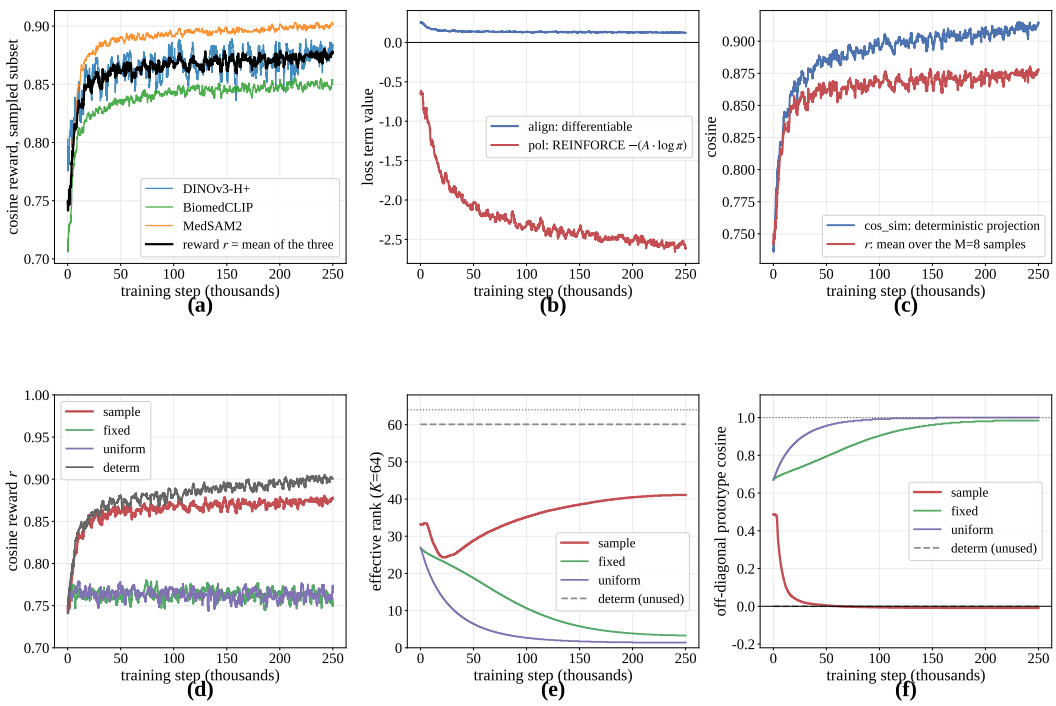

Figure 10: The OPD rollout policy as a policy gradient, measured over full training runs. **a)** Reward: each frozen expert's cosine verdict on the prototype subset the student itself sampled, and their mean. **b)** The two gradient channels: the differentiable alignment term (codebook, heads, bottleneck) and the codebook-GRPO term (gate logits only), plotted as *REINFORCE* in the legend. **c)** The sampled reward tracks the baseline's deterministic cosine, i.e. one reward definition serving both consumers. **d–f)** The same apparatus under four policy modes, matched on data, experts, $M{=}64$, $G{=}8$ and 250k steps; only policy mode differs. **d)** cosine reward r; **e)** codebook effective rank (dotted: M isotropic); **f)** mean off-diagonal prototype cosine ($+1$ = collapsed). Where the gate stops reading the position (*fixed*, *uniform*) the codebook collapses monotonically despite the entropy bonus, while *sample* stays diverse; the dashed *determ* curves in e,f) describe a bypassed, never-read tensor. All curves come from the training logs; these panels verify the mechanism and do not rank the modes.

spatial evidence overlaps only partially: the mean off-diagonal Spearman correlation is 0.46, with lower agreement for pancreatic lesions despite their high separability. Performance also decreases for smaller lesions and same-organ background, but remains well above chance. The teachers therefore provide informative yet non-redundant signals, motivating the pooled reward in Eq. 4 and the position-wise emphasis on difficult regions in Eq. 8.

The shared probe further shows that this lesion-related signal remains readable after distillation. In Fig. 4, HiPPO attains the highest probe AUC of the three student arms and lifts the student representation further into the range spanned by the frozen teachers, whereas the random-initialized control remains clearly below both trained models. The depth-resolved probe of Fig. 8, described above, locates this gain in the deeper encoder stages, where the method injects its supervision. Together, the two probes support transfer of the teachers' lesion-related structure without equating linear separability with segmentation performance.

The downstream ablations in Tab. 5(a) connect this representational evidence to task performance. No single visual expert matches the complete model, and removing any expert reduces macro AUC. DINOv3 illustrates the interaction clearly: its single-expert configuration reaches an AUC of 88.57%, yet removing it from the full pool still costs 2.41 points, indicating that its contribution depends on the other experts. Removing MedSAM2 reduces AUC from 90.33% to 87.61%, consistent with the importance of its boundary-oriented features. Separately, removing the privileged VLM branch reduces AUC and F1 by 1.56 and 2.12 points, respectively. Thus, the visual teachers contribute complementary spatial representations, while the VLM supplies an additional semantic signal outside the visual rollout reward.

B.4 SUMMARY: POLICY-BASED LEARNING

The rollout policy is measured twice: once as an optimization process, in Fig. 10, and once by what degrading it costs on task metrics, in Tab. 5b.

One reward definition, two gradient channels. Fig. 10a shows the reward actually consumed: each frozen expert's cosine verdict on the prototype subset the student itself sampled, together with their mean. Panel (b) separates the two paths that reward take, the differentiable alignment term, which reaches the codebook, the heads and the bottleneck, and the codebook-GRPO term, which reaches the gate logits only, so no gradient is routed through the frozen experts. Panel (c) shows the sampled reward tracking the deterministic cosine of the baseline over a full training run: the two consumers share one reward definition, and moving to sampling changes how the measured subset is selected, not what is being measured.

Selection, not stochasticity, prevents codebook collapse. Fig. 10d–f isolates the failure mode the entropy bonus of Eq. 6 exists to prevent. Because the four runs share data, experts, codebook and rollout budget ($M{=}64$ entries, $G{=}8$ rollouts per unit) and the 250k-step budget, the curves can differ only through how the gate reads the position. Where it does not, *fixed* and *uniform*, the codebook's effective rank falls monotonically and the mean off-diagonal prototype cosine climbs toward collapse even though the entropy term is active throughout: the bonus slows degeneration but cannot prevent it once the selection carries no position-specific information, because there is then no gradient that distinguishes one prototype from another. Under *sample* the codebook stays diverse, which is the mechanism claim of Sec. 3.2 stated as a measurement rather than as an assertion. The dashed *determ* curves describe a tensor that is never read and are included only to show that its geometry carries no information. These panels verify the mechanism; ranking the modes against each other on task metrics is what panel (b) of Tab. 5 does.

On AUC and BAcc the four non-learned variants order as deterministic projection $>$ input-independent dropout $>$ soft mixture $>$ uniform random rollout, every one below the learned policy. The degenerate variants are ordered by how much of the student state they still condition on, not by how much randomness they retain; had the gain been a stochastic regularization effect the ordering would be reversed. Deleting the codebook-GRPO term costs 4.11 and 2.40 points of F1 and macro Dice. Removing weight unification degrades AUC, BAcc, F1 and macro Dice by 1.51, 1.98, 2.04 and 2.07 points, which identifies the shared difficulty weight, rather than sampling alone, as what couples the rollout to the baseline distillation term.

Choice of the number of codebooks. $R{=}2$ reaches 89.34% AUC, 82.91% BAcc and 61.92% F1, and $R{=}3$ falls to 83.04% AUC and 75.17% BAcc, so $R{=}1$ is the setting we report. The macro-Dice figures quoted in this subsection (52.41% for the full model, 50.01% without codebook-GRPO, 50.34% without weight unification) are measured on the same runs as Tab. 5(a) but are not printed there, since that table reports the classification metrics only.

Scope of the claim. Fig. 10d–f reports one run per mode and verifies the mechanism; it does not rank the modes, which is what panel (b) of Tab. 5 does, on one benchmark. The two are complementary evidence for the same claim rather than independent replications of it.

B.5 SUMMARY: DYNAMIC SPATIAL WEIGHTING

Empirical behavior for dynamic weight. Fig. 11 evaluates the criteria and the realized weights by replaying the training loss of the checkpoint used for all main results on 40 training batches, i.e. 9,600 bottleneck positions per teacher. Two observations support the design. First, the arg-max shares realize the intended division of labor: $D_{k,p}$ is the active criterion on roughly two thirds to three quarters of positions, G_p on most of the remainder, and η_p on a sliver too thin to resolve in the stacked bars. This hierarchy is what a nearly converged student predicts: rollouts still lag the teachers on most positions, in-plane edges keep a teacher-independent floor of emphasis even where the features have already aligned, and the error term fires only at isolated mistakes, which the max in Eq. 10 lets override both salience and difficulty.

Second, the shape of the realized weight is the signature of the hinge, not of a rescaling: consistent with Eq. 14, the $\omega_{k,p}$ histograms peak well below their unit mean and carry a long right tail, while the $D_{k,p}$ histograms open with an atom at zero, the positions the hinge releases, before a broad hump

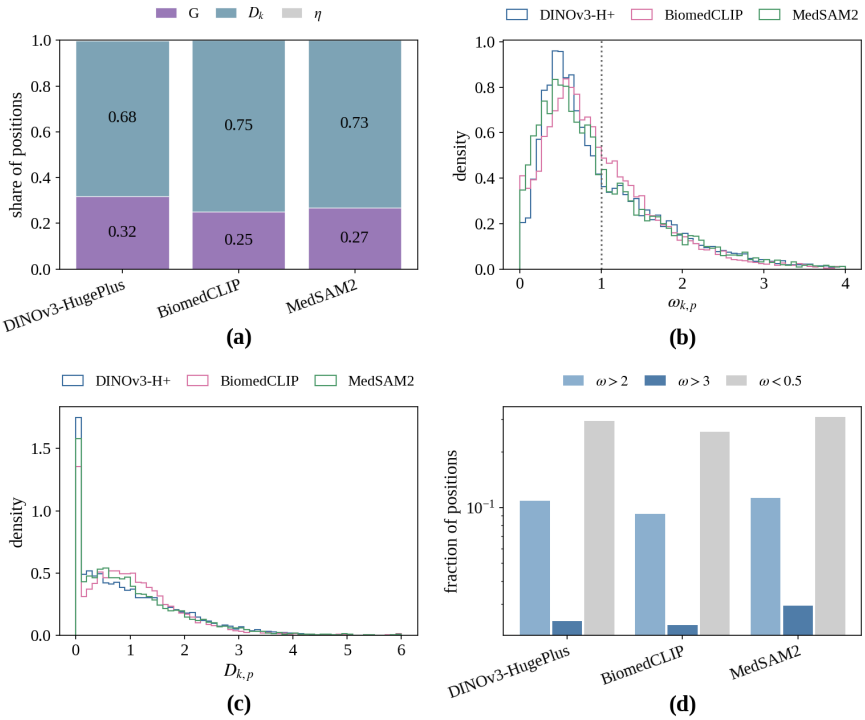

Figure 11: Statistics of the criteria and weights over 40 training batches. (a) share of positions at which each criterion attains the maximum in $u_{k,p}$; (b) per-teacher distributions of $\omega_{k,p}$; (c) per-teacher distributions of $D_{k,p}$; (d) tail fractions of $\omega_{k,p}$.

Table 11: Training cost on *MedBridge-CT.TumorExt*. Both runs share the same nnU-Net plan, batch size, schedule and a single GPU, so the difference is per-step compute only. Steady state excludes the first 20 epochs.

	nnU-Net	**HiPPO**	**Ratio**
Mean epoch time	199.6 s	390.2 s	1.955×
Steady-state epoch time (n=980)	199.2 s	390.7 s	1.961×
Total wall clock	55.44 h	108.40 h	+52.96 h

with a tail beyond four. The unit-mean constraint turns this into a budget: the roughly tenth of positions with $\omega_{k,p} > 2$ is funded by down-weighting the aligned majority, so emphasis is transferred to a small hard subset rather than amplified uniformly.

B.6 Training Cost

As shown in Tab. 11, under an identical nnU-Net configuration (same plan, batch size, schedule, and single GPU), HiPPO requires approximately 1.96× the per-epoch wall-clock time of the baseline nnU-Net, or 108.40 h versus 55.44 h in total. The difference comes from the additional forward and backward passes of the on-policy distillation and privileged-information modules, which are engaged only during training. At inference the student is the baseline nnU-Net itself, so the inference cost is unchanged.

C Supplementary Method Details

C.1 Student and Teacher Implementation

Student architecture. The shared encoder maps a 3D patch x to $f = E_\theta(x) \in \mathbb{R}^{C\times d\times h\times w}$ at the bottleneck resolution, and a dense lesion decoder follows the encoder. Both the patch shape and the per-axis stride product are taken from the nnU-Net (Isensee et al., 2021) plans of each dataset rather than fixed by us, so they are anisotropic and differ across datasets, and the bottleneck grid $|\Omega_n|$ differs with them; Tab. 13 lists the patch geometry selected for each benchmark. Where organ annotations are available, we use an organ-constrained variant that adds a second decoder for the organs: its logits are resized, encoded, and injected into each lesion-decoder stage without detaching, so the two branches shape each other, and an auxiliary organ segmentation loss is added to $\mathcal{L}_{\text{dense}}$. A multi-label presence head fuses lesion-decoder features across scales, applies adaptive max pooling, and predicts per-class presence. ϵ denotes a small fixed numerical floor throughout.

Teacher implementation. The pool comprises three visual teachers, DINOv3-H+, BiomedCLIP, and MedSAM2, and one medical VLM, HuluMed-4B, which supplies the privileged semantic target detailed in App. C.4. DINOv3-H+ (the HugePlus variant, also written DINOv3-HugePlus in figure labels) and HuluMed-4B are the specific checkpoints used in every experiment; the text refers to them simply as DINOv3 and HuluMed. All teachers are frozen and their targets enter as detached tensors. Visual teacher features are recomputed on the fly at every step rather than precomputed, since the pseudo-RGB views depend on the sampled patch; no feature cache is used. For MedSAM2, we take features from the last stage of its Hiera image encoder and bypass the memory-attention module, so no prompt or temporal context is used. BiomedCLIP patch tokens are used as provided; because this teacher was trained with image-level contrastive supervision, its dense targets are weaker a priori, consistent with its behavior in our teacher-pool study.

C.2 Visual-Feature Alignment

2.5D-to-3D construction. Each input channel is rendered separately. For a chosen axial slice, we build a pseudo-RGB image from that slice and its two immediate neighbors; a multi-channel crop thus yields one such view per channel, and the per-channel teacher features are averaged at the end. The input has already been intensity-standardized by the dataset-level normalization of the nnU-Net plans, so no HU windowing is applied at this stage.

The mapping to $[0, 1]$ depends on how the channel was standardized, since the two schemes place a lesion in very different parts of the intensity range. For a channel standardized with dataset-level statistics (wide spread, lesions near the mean), we clip the standardized values to a fixed symmetric window and map them affinely to $[0, 1]$. For a channel standardized case-by-case, a fixed window would flatten the lesion. Instead, we stretch it from a low percentile of its own histogram to its maximum, clipping to $[0, 1]$. Crucially, the three slices of a view share one window to preserve relative brightness. Each view is then normalized with a fixed channel mean and standard deviation, and bilinearly resized to the teacher's native input resolution.

Teacher k yields a token grid of size $C_k \times H_k \times W_k$ per slice, where H_k, W_k are determined by its native patch size. In-plane, we do not resample the grid: each bottleneck position takes the single token whose patch contains that position's center (nearest-neighbor correspondence). Along the axial axis, we spread several slices across the position's depth band, up to a fixed budget over the whole crop, and average their token features, matching a position against the slices it actually integrates. This produces $z_{k,p} \in \mathbb{R}^{C_k}$ at every $p \in \Omega$, where Ω indexes flattened batch and spatial positions. Because selecting one token per in-plane position and averaging across slices need not preserve the teacher's feature geometry, these aggregated targets serve as directional rather than metric references. The complete procedure is summarized in Algorithm 1.

C.3 Spatial Prioritization

Segmentation-error details. In the segmentation-error criterion $\eta_p = \mathbf{1}\{\hat{y}_p \neq y_p\}$, $\hat{y}_p$ is the student's arg-max class and y_p is the training label, both reduced to the bottleneck grid by nearest-

Algorithm 1: 2.5D-to-3D Feature Construction for Teacher k

Input : Standardized 3D crop $X \in \mathbb{R}^{C\times D\times H\times W}$, Teacher k with native resolution (H'_k, W'_k) and patch size defining grid (H_k, W_k), depth budget B.
Output: Aggregated target features $z_{k,p} \in \mathbb{R}^{C_k}$ for all $p \in \Omega$.

foreach *channel* $c \in \{1, \dots, C\}$ **do**
  foreach *axial slice* $d \in \{1, \dots, D\}$ **do**
    Stack slices $d-1, d, d+1$ to form pseudo-RGB view $V_{c,d}$;
    if *standardized via dataset-level statistics* **then**
      Clip $V_{c,d}$ to fixed symmetric window and map affinely to $[0, 1]$;
    else
      Compute low percentile P_{low} and max P_{max} of the view's histogram;
      Stretch $V_{c,d}$ from P_{low} to P_{max} and clip to $[0, 1]$ (shared across 3 slices);
    Normalize with fixed channel mean/std and resize to (H'_k, W'_k);
    Extract teacher token grid $T_{c,d} \in \mathbb{R}^{C_k\times H_k\times W_k}$;

foreach *spatial position* $p \in \Omega$ **do**
  Find in-plane nearest-neighbor token index (i_p, j_p) corresponding to p's center;
  Identify axial depth band $\mathcal{S}_p$ around p, bounded by budget B;
  Average tokens across depth band: $z_{k,p} \leftarrow \frac{1}{|\mathcal{S}_p|}\sum_{d\in\mathcal{S}_p} T_{c,d}[:, i_p, j_p]$;

return $\{z_{k,p}\}_{p\in\Omega}$ (used as directional references);

neighbor interpolation. The criterion is recomputed every step, so it tracks the student's current errors rather than a fixed difficulty map, and it is detached like the other two criteria.

Moment interpretation of the difficulty criterion. To isolate the alignment-difficulty component, consider its linearized form before the hinge and clipping and without the stabilizing ϵ:

$$D^{\text{lin}}_{k,p} = t + \frac{\bar{\delta}_{k,p} - \mu_{\bar{\delta}_k}}{\sigma_{\bar{\delta}_k}}, \tag{12}$$

where $\bar{\delta}_{k,p} = 1 - \bar{q}_{k,p}$ and $\sigma_{\bar{\delta}_k} > 0$. We temporarily exclude the maximum with the other criteria. Let μ_{δ_k} and σ_{δ_k} denote the empirical mean and standard deviation of the deterministic matching distance $\delta_{k,p}$ over Ω. All empirical moments in this derivation use normalization by $1/|\Omega|$. Since the standardized term in Eq. 12 is centered,

$$\frac{1}{|\Omega|}\sum_{p\in\Omega} D^{\text{lin}}_{k,p} = t. \tag{13}$$

The corresponding unit-mean-normalized criterion satisfies

$$\begin{aligned}\frac{\sum_{p\in\Omega} D^{\text{lin}}_{k,p}\delta_{k,p}}{\sum_{p\in\Omega} D^{\text{lin}}_{k,p}} &= \mu_{\delta_k} + \frac{\text{Cov}_{p\in\Omega}\big(\bar{\delta}_{k,p}, \delta_{k,p}\big)}{t\,\sigma_{\bar{\delta}_k}} \\ &= \mu_{\delta_k} + \frac{\rho_k}{t}\sigma_{\delta_k}.\end{aligned} \tag{14}$$

Here, the second equality holds when $\sigma_{\delta_k} > 0$, and ρ_k is the empirical Pearson correlation between $\bar{\delta}_{k,\cdot}$ and $\delta_{k,\cdot}$. If $\sigma_{\delta_k} = 0$, the covariance correction vanishes and the first equality remains well-defined.

Eq. 14 makes the role of the rollout signal explicit. Positive correlation between the rollout-derived and deterministic distances raises the normalized criterion above the unweighted mean, with an increment proportional to $\rho_k\sigma_{\delta_k}/t$. When the two quantities are weakly correlated, the additional emphasis correspondingly diminishes.

In the special case $\rho_k = 1$, Eq. 14 reduces to $\mu_{\delta_k} + \sigma_{\delta_k}/t$. For context, let $\text{CVaR}^{\text{upper}}_{\alpha}$ denote the mean over the worst α fraction of the empirical distance distribution. The distribution-free moment bound is

$$\text{CVaR}^{\text{upper}}_{\alpha}[\delta_k] \le \mu_{\delta_k} + \sigma_{\delta_k}\sqrt{\frac{1-\alpha}{\alpha}}. \tag{15}$$

The standard-deviation coefficients in Eqs. 14 and 15 coincide when

$$\alpha = \frac{t^2}{1+t^2}. \tag{16}$$

This links the scale parameter t to a familiar tail-sensitive coefficient when the rollout-derived and deterministic distances are strongly aligned. The complete objective retains this motivation while incorporating the bounded nonlinearities and additional weighting signals described below.

In the deployed criterion, the hinge ensures non-negative weights and suppresses positions whose rollout-derived distance is sufficiently below the batch mean, while clipping limits the influence of outliers. Because the resulting weights are detached, these operations redistribute gradient magnitude across positions rather than introduce a differentiable quadratic penalty. The deployed weight further incorporates transition salience and segmentation error through the maximum, followed by batchwise normalization. Eq. 14 therefore characterizes the alignment-difficulty component that motivates the complete feature-distillation loss.

Interaction among criteria. The three criteria interact asymmetrically. With $t = 1$ and a positive clipping bound, $D_{k,p} > 1$ exactly when the rollout-derived distance is above its batch mean. At such positions, it necessarily dominates $G_p, \eta_p \in [0, 1]$. At positions with at-or-below-mean rollout difficulty, $D_{k,p} \leq 1$, allowing transition salience or segmentation error to retain structurally important or currently misclassified regions that would otherwise receive little alignment-based weight. The maximum therefore prioritizes above-mean alignment difficulty while preserving independent routes for spatial structure and task error.

Per-teacher standardization largely removes location and positive-scale differences in each teacher's rollout-derived distances. It does not make teacher contributions identical: after the hinge, clipping, maximum, and batchwise normalization, the realized contribution of a teacher is not determined solely by $\mu_{\delta_k} + \rho_k \sigma_{\delta_k}/t$. Because the statistics and normalizer are recomputed from the current batch, the weight assigned to a position also adapts to the other positions in that batch.

C.4 Privileged Semantic Branch

Privileged-view construction. We take the connected component of the lesion mask with the largest volume, treating components below a few voxels as no lesion at all, and enlarge its bounding box by a fixed relative margin to retain perilesional context. The cropping window is then clamped so that the achieved magnification, taken as the geometric mean over the three axes, lies within a fixed range: the lower bound keeps a small finding large enough to describe, the upper bound keeps enough surrounding tissue for the teacher to localize it. The window is resampled trilinearly back to the original crop shape and rendered into the same axial-slice stack the VLM receives on the unprivileged path, so the two branches differ in magnification and annotation-based localization only, not in image format. The teacher branch receives this view and no soft tokens; the student branch receives the slices of its own crop together with its N_s soft tokens. For lesion-free crops there is no component to magnify, so the token term is skipped and only the readout term is applied, with an all-negative target.

Token matching details. The teacher produces a greedy rollout of up to T tokens, and the student branch is scored on the same prefix. At each scored position both distributions are restricted to the teacher's Top-K_{tok} tokens and renormalized over that set before the divergence of Eq. 11 is taken. This is a restriction rather than a coarse-graining: no residual bin collects the discarded mass, so the term scores agreement on the teacher's retained support and by itself carries no penalty for student mass placed outside it. The support is the teacher's own, which is what keeps the term informative where the teacher is confident. We use JSD rather than a directed KL because it remains bounded when the student initially assigns near-zero mass to a teacher-preferred token. Scoring the student on the teacher rollout places the objective in the SeqKD (Kim & Rush, 2016) or ImitKD family (Lin et al., 2020), rather than the student-sampled family (Agarwal et al., 2024; Gu et al., 2024).

Soft-token projection. Ψ is a single linear map from the bottleneck width to the VLM hidden size followed by a layer normalization, randomly initialized and trained from scratch with the student; it is not warm-started from the VLM's own image projector. A learned per-slot embedding of the

same dimension is added to the N_s tokens so that slots remain distinguishable to the frozen model beyond their content.

C.5 Optimization and Gradient Paths

Tab. 12 enumerates which parameters each objective updates and whether the gradient reaches the shared encoder under the hard Bernoulli sampling of Eq. 3.

Table 12: Gradient paths under the hard Bernoulli sampling in Eq. 3, without a straight-through estimator. E_θ: encoder; π_ϕ: sampling policy; $\mathcal{E}$: unified codebook; h_k: teacher-specific projectors; Ψ: soft-token projection.

Objective	**Updates**	**Reaches E_θ**	**Signal**
$\mathcal{L}_{\text{dense}}$	E_θ, decoders	yes	dense, differentiable
$\mathcal{L}_{\text{cls}}$	E_θ, presence head	yes	view-level, differentiable
$\mathcal{L}_{\text{policy}}$	π_ϕ, $\mathcal{E}$, $\{h_k\}$	via π_ϕ	rollout, teacher-relative advantage
$\mathcal{L}_{\text{feat}}$	E_θ, $\{h_k\}$	yes	dense, priorities detached
$\mathcal{L}_{\text{tok}}$, $\mathcal{L}_{\text{readout}}$	E_θ, Ψ	yes	view-level, through frozen VLM

Exactly one term in the objective is non-differentiable: the codebook-GRPO term of Eq. 6, whose gradient is supplied by the score-function estimator. We use no straight-through estimator and no continuous relaxation of the Bernoulli action. The alignment term of the same equation therefore reaches only $\mathcal{E}$ and $\{h_k\}$, because the sampled composition is formed from the codebook with a detached selection, and it is this term that determines what the prototypes encode. Both $\mathcal{L}_{\text{feat}}$ and the semantic objectives are differentiable all the way to E_θ: the former acts on the deterministic representation f through $\{h_k\}$, the latter on f pooled into soft tokens through Ψ.

The semantic term is applied every q steps to reduce training cost, and all trainable modules, including Ψ, share one optimizer. The intermittent schedule is therefore not exactly equivalent to applying α_{vl}/q at every step: the two agree in the long-run average gradient they deliver but differ in the within-window sequence the momentum buffer accumulates. The method hyper-parameters are broadly consistent across benchmarks rather than tuned per dataset, the clearest exception being λ_{cls}, which weights the presence head; that head is trained only where organ annotations exist and is switched off where they do not, so λ_{cls} follows the dataset. The values we use are collected in Tab. 14, and the per-benchmark geometry, batch and schedule in Tab. 13. α_{vl} scales the semantic objective against the segmentation losses present in that run; it is varied only in the mechanism ablations, where setting it to zero is precisely how the semantic term is removed. The training length is likewise set by the budget of each comparison rather than shared, so it differs between the CT, MR and PET/CT settings.

The empty action. The action in Eq. 3 consists of M independent Bernoulli variables, and we intentionally allow the empty subset $a_{i,u} \equiv 0$, which yields $\tilde{f}_{i,u} = 0$. This case does not require rejection sampling or special handling. Because each projector h_k is an affine transformation followed by layer normalization, $h_k(0)$ is a fixed, finite teacher-specific output. Moreover, q_k in Eq. 7 uses stabilized normalization, so the resulting reward remains finite even in this case. The empty rollout therefore participates in the same group-relative ranking as the other $G - 1$ candidates and is optimized by the same objective. We additionally bound the pre-sigmoid logits and apply entropy regularization to keep the Bernoulli policy away from numerically degenerate probabilities and to discourage premature collapse. Importantly, the validity of the method does not rely on the empty action having negligible probability; it is simply one admissible codebook configuration handled by the general formulation.

D Implementation Details

All experiments are conducted on NVIDIA H20 GPUs (96 GB). Our method and the nnU-Net-family comparators inherit their configuration from nnU-Net's automatic planner for the dataset at hand: patch size, target spacing, network depth and channel widths, intensity normalization statistics, and the optimizer recipe of SGD, an initial learning rate of 10^{-2} under polynomial decay, deep supervision and foreground oversampling. Training runs in bf16 automatic mixed precision, and the

Table 13: Per-benchmark training configuration of HiPPO. Patch and spacing are $D\times H\times W$ as selected by the nnU-Net planner, except in *MedBridge-CT.Nodule* †, which trains on nodule-centered CT blocks of a fixed shape at their acquisition resolution. Rows marked ‡ run one epoch as a full pass over the training loader. Batch size is per GPU.

Benchmark	Input	Patch	Spacing (mm)	Batch	Epochs × steps	λ_{cls}	GPUs
M.B.-CT.TumorExt	CT	$80\times128\times192$	$2.50\times0.80\times0.81$	2	1000×250	0.5	1
M.B.-CT.TumorInt	CT	$96\times160\times160$	$3.00\times1.00\times1.00$	2	1000×250	0.5	4
M.B.-CT.Nodule	CT blocks†	$64\times128\times128$	$2.00\times0.59\times0.59$	4	$50^{\ddagger}$	1.0	4
M.B.-MR	DWI	$80\times96\times80$	$2.00\times2.00\times2.00$	2	1000×250	0.5	1
M.B.-PET/CT-Seg	CT + PSMA	$112\times192\times112$	$3.82\times3.27\times3.82$	2	1000×250	0.3	1
M.B.-PET/CT-Cls	CT + FDG	$112\times160\times128$	$3.00\times2.04\times2.04$	2	1000×250	0.5	4

Table 14: Main hyper-parameters of HiPPO, in three panels: objective weights, rollout and prioritization, and the privileged semantic branch. These values are broadly consistent across benchmarks rather than tuned per dataset; the per-benchmark geometry, batch, schedule and λ_{cls} are given in Tab. 13. Optimizer, learning-rate schedule and precision follow the nnU-Net recipe described in the text. Each codebook entry is $e_m \in \mathbb{R}^{256}$.

Objective weights		Rollout and prioritization		Semantic branch	
λ_{dis} (distillation)	0.5	M (codebook entries)	64	T (rollout length)	256 tokens
$\lambda_{\mathrm{policy}}$ (rollout term)	0.5	G (rollouts per unit)	8	K_{tok} (Top-K_{tok})	100
λ_{sup} (readout)	0.5	t (target agreement)	1	N_s (soft tokens)	8
α_{vl} (semantic term)	0.1	b (outlier bound)	5	Magnification	$[2\times, 4\times]$
β (entropy bonus)	0.01	q (semantic interval)	10 steps		
λ_{cls} (presence)	per benchmark	Slices per crop	32	R (codebooks)	1

three visual experts and the VLM are kept frozen throughout, so the trainable set is the student, the codebook head and Ψ. The MONAI-based comparators keep their authors' defaults for architecture, loss, optimizer and augmentation. Their target spacing and patch size are also set per dataset, but only as far as each architecture allows: the crop has to be divisible by the encoder's window and patch sizes and the transformer-based comparators are tuned for near-isotropic crops, so their geometry stays close to the planner's choice without being equal to it. What is held fixed across frameworks is the schedule. Every method is trained for the same number of epochs, and HiPPO is never given more crops per optimizer step than a comparator. Counted in samples seen, HiPPO is therefore at or below every method it is compared with. Each comparator is thus evaluated under the configuration its own framework recommends.

Tab. 13 gives the per-benchmark configuration of HiPPO. The five settings that go through the nnU-Net pipeline share the budget above and differ only in what the planner selects for each dataset, in λ_{cls}, and in how many GPUs the global batch is spread over; the batch size is the per-GPU value the trainer uses, which is 2 on every single-GPU run irrespective of the batch the planner proposed. The remaining method hyper-parameters are broadly the same across settings and are listed in Tab. 14. *MedBridge-CT.Nodule* is the one exception to the recipe above: it does not use planner-selected patches but nodule-centered CT blocks at their acquisition resolution, and it is trained for 50 epochs, so one epoch is a full pass over the training loader.

E Detailed Discussion for Related Work

E.1 Medical Foundation Models

Foundation models encode prior knowledge at a scale difficult to obtain from any single task-specific medical cohort and span several complementary families. In 2D, biomedical vision–language models such as BiomedCLIP (Zhang et al., 2023) align large collections of biomedical image–text pairs in a shared embedding space, providing cross-modal representations that can transfer to unseen pathologies and downstream tasks. General self-supervised backbones such as DINOv2 (Oquab et al., 2024) and DINOv3 (Siméoni et al., 2025), by contrast, learn semantically rich visual features

from large-scale image corpora without text supervision. Recent work further indicates that medical image encoders pretrained with vision-only objectives can match or exceed vision–language pretraining on downstream segmentation tasks (Pérez-García et al., 2025), suggesting that dense visual prediction can benefit substantially from representations learned without language alignment. Promptable segmentation models, including the SAM series (Kirillov et al., 2023; Ravi et al., 2025; Carion et al., 2026) and their medical adaptations (Ma et al., 2024a; 2025; Zhu et al., 2024; Liu et al., 2025), provide a different form of prior knowledge. They accept spatial prompts such as points, boxes, or scribbles and produce spatially resolved masks across diverse anatomical structures and imaging conditions. Their dense feature maps and refined mask representations consequently provide localization and boundary information.

In 3D, self-supervised methods such as VoCo (Wu et al., 2024) and UniMiSS+ (Xie et al., 2024) learn volumetric context directly from unlabeled scans, while SuPreM (Li et al., 2024) transfers representations pretrained on large-scale automatically annotated abdominal CT. All three capture inter-slice relationships, anatomical structure, and organ topology that cannot be represented completely by independently processed 2D images. Complementarily, 3D CT vision–language backbones such as Merlin (Blankemeier et al., 2026), M3D (Bai et al., 2024), and RadFM (Wu et al., 2025) associate volumetric images with radiology reports or structured clinical records. Related systems for volumetric diagnosis and report generation (Hamamci et al., 2024; Bassi et al., 2025) further ground image findings in clinical language.

At a broader semantic level, medical multimodal LLMs such as LLaVA-Med (Li et al., 2023), Lingshu (Xu et al., 2025), HuluMed (Jiang et al., 2025), and ClinFusion (Yuan et al., 2026a) combine image understanding with visual question answering, report generation, and clinically oriented semantic reasoning. These models can provide diagnostic semantics that are difficult to recover from dense visual features alone.

Taken together, these model families provide complementary but non-commensurate forms of supervision: structural and appearance cues from self-supervised encoders, localization and boundary information from segmentation models, cross-modal alignment from vision–language models, and diagnostic semantics from medical multimodal LLMs. Their architectures, objectives, spatial resolutions, and representation spaces differ substantially, however, making direct feature matching or fixed aggregation inadequate in general. We therefore treat them as a heterogeneous pool of frozen teachers rather than as interchangeable backbone alternatives.

E.2 On-Policy Distillation

On-policy distillation broadly refers to training a student on outputs induced by its own evolving policy, with teacher feedback evaluated on those outputs rather than exclusively on teacher-generated data. In classical sequential imitation learning, querying the teacher along learner-induced trajectories can reduce compounding error relative to purely off-policy supervision, from quadratic to linear growth in the sequence horizon under the corresponding assumptions (Ross et al., 2011). This analysis motivates modern on-policy distillation in autoregressive models, although its sequential guarantee does not directly transfer to the dense representation setting considered in this work.

Modern language-model distillation commonly matches teacher and student predictive distributions along student-generated sequences rather than teacher-generated ones (Agarwal et al., 2024; Gu et al., 2024; Ko et al., 2024). This development builds on earlier sequence-level and imitation-based distillation frameworks (Kim & Rush, 2016; Lin et al., 2020), which already recognized the exposure bias introduced when a student is trained only on trajectories that it would not generate itself. Related group-relative policy objectives normalize the rewards of multiple sampled outputs against their group statistics, avoiding the need for a separately trained value model. GRPO, for example, replaces a learned per-token critic with a group-relative baseline (Shao et al., 2024), and variants of this design have become increasingly common in large-model post-training (Yang et al., 2025; Zhao et al., 2026).

The on-policy principle has recently been extended to vision–language students. VOLD (Bousselham et al., 2026) transfers reasoning knowledge from text-only LLMs to VLMs by distilling teacher supervision along visual-question rollouts. Visual-Advantage OPD (VA-OPD) (Liu et al., 2026) identifies vision-dependent tokens through a teacher-derived visual advantage that contrasts predictions under the original and resolution-degraded images. The resulting signal redistributes su-

pervision across tokens and rollouts, concentrating it on predictions that depend on visual evidence rather than language priors. Med-OPD adapts evidence-aware weighting to medical VLMs and uses ground-truth answer hints as privileged signals to steer the teacher toward clinically relevant tokens (Qian et al., 2026). CA-OPD extends the paradigm to structured visual prediction problems such as GUI grounding and OCR, where the student autoregressively emits structured token sequences (Li et al., 2026a). It uses teacher confidence to identify unreliable student transitions under a strict-to-relaxed correction schedule: corrected positions receive direct cross-entropy supervision, while retained positions are trained using the teacher's predictive distribution.

A parallel line of work introduces on-policy self-distillation through controlled asymmetries between two instances of the same model. Vision-OPD (Yuan et al., 2026b), for example, uses a single multimodal model as both a crop-conditioned teacher and a full-image student. The two branches share model parameters but receive different views, allowing the student to recover details visible to the teacher under magnification. S2VOPD (Li et al., 2026c) removes the privileged teacher view and instead degrades the student's input through downscaling and noise, creating an informative teacher–student asymmetry without additional annotations. OPD-V (Bi et al., 2026) addresses modality imbalance and introduces a modality-balance trust region that selects suitable on-policy tokens for self-distillation according to modality-balance logit margins.

Despite their differences, the methods discussed above largely instantiate on-policy learning in an autoregressive output space. Their actions are tokens or structured output tokens, and teacher feedback is defined over generated tokens or token sequences. Their targets are correspondingly visual-question-answering, reasoning, or structured-prediction outputs rather than dense spatial representations. A language-free dense visual predictor has neither a vocabulary-valued action space nor a text decoder on which token-level feedback can be imposed. Moreover, heterogeneous visual teachers produce representations with different dimensions, spatial resolutions, and semantic geometries. Extending the on-policy principle to this setting therefore requires (i) a representation-level action space, (ii) a mechanism for constructing student-conditioned candidate representations, and (iii) teacher-specific compatibility functions that can evaluate those candidates without forcing all teachers into a common representation space. Our formulation addresses this gap by treating codebook compositions as student-conditioned latent candidates and evaluating each candidate against the applicable dense targets in the native space of each teacher.

E.3 Privileged Information Distillation

Learning using privileged information (LUPI) considers a teacher that has access during training to side information unavailable to the deployed student (Vapnik & Vashist, 2009; Vapnik & Izmailov, 2015). This asymmetry naturally connects to knowledge distillation, where privileged information can improve the teacher's target or representation without changing the student's inference-time inputs (Lopez-Paz et al., 2015; Pechyony & Vapnik, 2010).

In computer vision, privileged information has taken several forms, including high-resolution images for efficient super-resolution (Lee et al., 2020), future frames for online action detection (Zhao et al., 2022), and longitudinal screening histories for mammography risk prediction (Karimian et al., 2026). In medical imaging, anatomy-privileged distillation uses organ and tumor masks available only during training to guide a teacher's token routing over 3D MRI volumes (Go et al., 2026). The resulting spatial routing knowledge is distilled into a mask-free student for perineural invasion prediction, demonstrating that annotation-derived spatial privilege can benefit a model even when the corresponding annotation is absent at inference.

Privileged supervision is complementary to on-policy distillation. Several multimodal OPD methods already condition teacher supervision on information unavailable to the deployed student, such as a ground-truth answer hint (Qian et al., 2026) or a magnified evidence view (Yuan et al., 2026b). These methods nevertheless remain text-decoder VLMs: their rollouts and supervision are defined over autoregressive token sequences, and the privileged information modifies the teacher's token-level feedback. In dense medical vision, spatial annotations can instead define localized views, regions of interest, or routing signals that are available only during training. Anatomy-privileged distillation (Go et al., 2026) demonstrates the usefulness of such spatial information but transfers a fixed routing target rather than evaluating student-conditioned sampled representations.

In our setting, privileged information and on-policy learning enter through two complementary branches. Frozen visual teachers evaluate student-conditioned, per-position latent rollouts constructed from non-privileged image views. Separately, a frozen medical VLM observes an annotation-derived lesion-centered view and provides a differentiable semantic target; it does not participate in the per-position rollout reward. The same language-free volumetric student therefore benefits from two distinct training-only asymmetries: adaptive teacher feedback on its own sampled latent representations and semantic supervision derived from information unavailable at deployment. To our knowledge, prior work has not combined representation-level on-policy transfer from heterogeneous dense visual teachers with a privileged semantic distillation branch in a compact, language-free visual predictor. This combination is central to our setting: the on-policy branch determines which candidate visual representations are useful under the student's current state, whereas the privileged branch transfers lesion-centered diagnostic semantics that dense visual agreement alone does not provide.

F Limitations and Future Work

Several limitations define the scope of the present study. The privileged semantic branch requires lesion-centered views during training. Where localization annotations are unavailable even in the training data, only the visual-teacher pool remains applicable; quantifying performance in this setting is left to future work. In addition, our experiments use one particular combination of frozen teachers, although the framework can accommodate other teacher pools through teacher-specific scoring functions. The current evaluation is also restricted to 3D medical images, and its applicability to 2D images and video has not yet been established.

Future work will explore larger and modality-specific teacher pools, expand the smallest evaluation cohorts, and develop a unified, dimension-adaptive framework for heterogeneous spatial and temporal representations. Such a framework would extend the current 3D-specific formulation to broader medical imaging modalities, including 2D images and video, while preserving shared mechanisms for representation-level actions, teacher-specific feedback, and privileged knowledge transfer.