WHOLESIGHT: LEARNING TO THINK WITH WHOLE SLIDE IMAGES

Anonymous authors
Paper under double-blind review

ABSTRACT

Reasoning over a whole slide image (WSI) requires collecting evidence from sparse tissue regions across spatial locations and magnifications. A WSI agent must decide where to look, what to inspect at higher magnification, and how new observations should guide subsequent reasoning. Existing WSI multimodal systems have improved either slide representations or active inspection, but visual evidence acquisition is typically fixed, prescribed at inference time, or learned separately from reasoning. We introduce **WholeSight**, which instead treats the WSI as an interactive environment in which visual search and reasoning are learned jointly as a single policy. A WSI harness exposes regional search and local inspection as executable actions, allowing the agent to acquire and interpret evidence over successive observations. Supervised fine-tuning provides a cold start for reasoning and tool use from complete interaction trajectories, after which reinforcement learning optimizes evidence acquisition and reasoning through the agent's own slide interactions. Across 2,956 questions spanning ten evaluation settings, WholeSight ranks first in eight and outperforms the strongest evaluated WSI system of comparable scale by 11.10 percentage points overall. Relative to its original multimodal large language model backbone under the same visual interface, the learned policy improves performance across every evaluation, including slide sources outside the training collection drawn from The Cancer Genome Atlas. Our code is available at https://anonymous.4open.science/r/anonymous-method-release-A53C/.

1 INTRODUCTION

A pathology whole slide image (WSI) can contain billions of pixels, while diagnostically relevant evidence may occupy only sparse regions across the slide. Low magnification reveals tissue architecture, whereas high magnification exposes cellular and morphological detail. Answering a question may therefore require integrating evidence from multiple spatially separated regions at different scales. Reasoning over a WSI involves not only interpreting visual evidence, but also deciding how to acquire it. The model must determine where to look, what to inspect at higher magnification, and how each new observation should guide subsequent reasoning. Recent WSI agents and benchmarks increasingly require models to locate relevant regions, inspect them at appropriate magnification, and reason over the resulting observations (Liao et al., 2026; Chen et al., 2026a; Yang et al., 2026).

Slide-level and patch-level foundation models organize gigapixel information into representations suitable for downstream reasoning (Chen et al., 2022; Xu et al., 2024; Ding et al., 2025), while slide-level multimodal large language models (MLLMs) connect these representations to language (Chen et al., 2025b; Liang et al., 2025). Interactive WSI agents acquire additional slide observations through predefined workflows at inference time (Buckley et al., 2026; Chen et al., 2026a; Yang et al., 2026). Other approaches learn inspection or navigation components separately from the MLLM responsible for downstream reasoning (Chen et al., 2026b; Huang et al., 2026). These systems improve gigapixel slide representation and access, but visual evidence acquisition and reasoning are still typically specified or optimized separately. It remains unclear whether visual search and pathology reasoning can instead be learned jointly as a generalizable policy through interaction with the WSI.

We introduce **WholeSight**, which treats the WSI itself as an interactive training environment for learning visual search and pathology reasoning jointly as a single policy (Figure 1). The policy requests tissue observations, interprets them in the accumulated visual context, and decides what to

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.

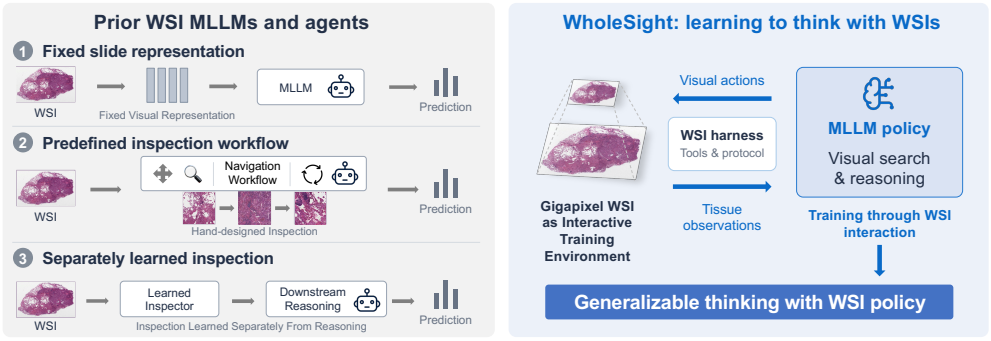

Figure 1: **Learning to think with WSIs.** Fixed slide representations, predefined workflows, and separately trained inspection components organize visual access in prior WSI multimodal systems. WholeSight learns visual search and reasoning through interaction with the WSI. The WSI harness executes visual actions and returns tissue observations.

inspect next or when to answer. Each action changes the evidence available for subsequent reasoning. We use *thinking with WSIs* to describe this process in which visual observations are actively acquired and used during reasoning (Zheng et al., 2026; Su et al., 2025b;a).

To make this interaction executable at gigapixel scale, we build a WSI harness that exposes multiscale visual access as actions over the slide pyramid. Regional search retrieves $5\times$ coarse views of candidate tissue, while local inspection returns spatially linked $20\times$ observations selected by the policy. The harness defines the available visual operations, while the policy learns how to use them through interaction. The same interface is used for supervised fine-tuning (SFT) trajectory collection, online reinforcement learning (RL), and evaluation (Jin et al., 2025; Feng et al., 2026; Zheng et al., 2026; Su et al., 2025b; Lu et al., 2026).

Complete tool-use interaction trajectories provide an SFT cold start, followed by online RL through the policy's own slide interactions. We evaluate the resulting 9B model on 2,956 questions across ten settings spanning recognition, diagnosis, grading, metastasis assessment, and detailed pathology questions. WholeSight ranks first in eight of the ten evaluations and exceeds the strongest evaluated WSI system of comparable scale by 11.10 percentage points overall. Under the same visual interface, it improves over the original Qwen3.5-9B backbone by 12.01 points, with gains of 11.86 points on slides from outside The Cancer Genome Atlas (TCGA) and 12.36 points on TCGA slides. Together, these results show that WSI interaction enables joint learning of visual evidence acquisition and pathology reasoning, supporting generalizable agentic reasoning over gigapixel pathology images.

2 Related Work

Pathology representations and multimodal reasoning models. Multiple-instance learning (MIL) models with attention, CLAM, and TransMIL learn to aggregate tissue instances (Ilse et al., 2018; Lu et al., 2021; Shao et al., 2021); HIPT and Prov-GigaPath organize information hierarchically and across whole slides (Chen et al., 2022; Xu et al., 2024). UNI and Virchow provide pathology feature encoders (Chen et al., 2024; Vorontsov et al., 2024), while PLIP, CONCH, KEEP, and TITAN connect visual representations to language (Huang et al., 2023; Lu et al., 2024a; Zhou et al., 2026; Ding et al., 2025). Quilt-1M provides paired pathology images and text. Assistants trained to follow instructions, including Quilt-LLaVA and PathChat, support interpreting tissue in language (Ikezogwo et al., 2023; Seyfioglu et al., 2024; Lu et al., 2024b). SlideChat and WSI-LLaVA extend multimodal modeling to whole slides (Chen et al., 2025b; Liang et al., 2025). At the patch level, Patho-R1, PathReasoner-R1, and PathScale-R1 receive additional training in pathology reasoning while operating on supplied images through visual encoders for general use. They take small patch images as input rather than slide level features (Zhang et al., 2026; Jiang et al., 2026a; Phan et al., 2026).

WSI access and agentic reasoning. GIANT, PathAgent, and PathNavigate organize iterative WSI exploration through interaction workflows (Buckley et al., 2026; Chen et al., 2026a; Yang et al., 2026).

SlideSeek uses multiple agents for hierarchical diagnostic reasoning over whole slides (Weishaupt et al., 2026), while BEACON selects evidence by expected information gain under competing diagnostic hypotheses (Wong et al., 2026). HistoSelect selects tissue according to the task (Huang et al., 2026). CPathAgent trains staged multiscale pathology analysis (SUN et al., 2025); Pathology-CoT learns from expert viewing behavior (Wang et al., 2026); AdaptivePath trains a Navigator while downstream interpretation and deliberation remain separate (Chen et al., 2026b).

Visual policy learning with tools. ReAct and Toolformer connect language generation to executable actions (Yao et al., 2023; Schick et al., 2023), and Search-R1 and ReTool optimize tool use with execution feedback (Jin et al., 2025; Feng et al., 2026). V* and Visual Sketchpad develop visual search and visual computation interfaces (Wu & Xie, 2023; Hu et al., 2024). DeepEyes, Pixel Reasoner, OpenThinkIMG, and Visual-ARFT train visual reasoning policies that invoke image operations during reasoning (Zheng et al., 2026; Su et al., 2025a;b; Liu et al., 2025); DeepEyesV2 extends this setting to adaptive use of image operations, code execution, and web search within one multimodal reasoning loop (Hong et al., 2026). In medical imaging, MedVR trains an end-to-end agentic visual reasoning policy with image manipulation tools, while MedVistaGym studies training with tools across medical image tasks (Jiang et al., 2026b; Lu et al., 2026). WholeSight trains visual search and pathology reasoning together through interaction with gigapixel WSIs.

3 METHODOLOGY

3.1 WSI AS AN INTERACTIVE TRAINING ENVIRONMENT

Let S denote a WSI and q a pathology question. At interaction step t the policy receives the question, the observations acquired so far, and the preceding model turns,

$$c_t = \mathcal{C}_\rho(q, y_{<t}, o_{\leq t}), \qquad (r_t, a_t) \sim \pi_\theta(\cdot \mid c_t), \tag{1}$$

where each preceding turn is $y_j = (r_j, a_j)$, r_t is generated reasoning, and a_t is either a visual request or a terminal answer. The context function $\mathcal{C}_\rho$ determines which acquired images are available to the model. A visual request is executed against the slide,

$$(s_{t+1}, o_{t+1}) = \mathcal{E}(S, s_t, a_t), \tag{2}$$

where s_t denotes the current interaction state and $\mathcal{E}$ denotes the WSI environment. The WSI harness implements this transition by executing a_t on the slide and returning the tissue observation o_{t+1}.

The action at step t determines the tissue observation at step $t + 1$, which the policy uses for its next reasoning step and visual request. During RL, the policy generates trajectories through these interactions and is updated by rewards from their task outcomes.

3.2 MULTISCALE EVIDENCE ACQUISITION

The policy accesses the WSI environment through two visual actions (Figure 2). `Search` implements regional search, which takes pathology keywords generated by the model and the requested number of views. The WSI harness selects and ranks candidate regions by similarity between keywords and $20\times$ patches in each region, then returns one $5\times$ coarse view of each selected region. The policy compares tissue architecture across these views and selects a coarse view identifier and a local bounding box for `Crop_view`, which provides an interface for local inspection. The WSI harness maps this selection to the slide and returns a $20\times$ detailed observation. Each detailed observation remains tied to the coarse view and location from which it was cropped. The similarity computation backend of `Search` remains fixed during training, and the policy learns to formulate `Search` requests and act on the returned observations. Details of the tools are described in Appendix A.

Each episode starts with one `Search` action and can contain zero or more `Crop_view` actions before the terminal answer. A local inspection can revisit another location within the same coarse view or move to a different coarse view. After each successful action the new image enters the next model turn. The policy can then request another detailed observation or emit the terminal answer.

By default, we retain all images returned in the context of both training and evaluation. Each successful `Crop_view` appends a detailed observation to the visual history. Later turns can compare the original coarse views with earlier collected observations. The action history and returned images preserve their temporal order with the model reasoning.

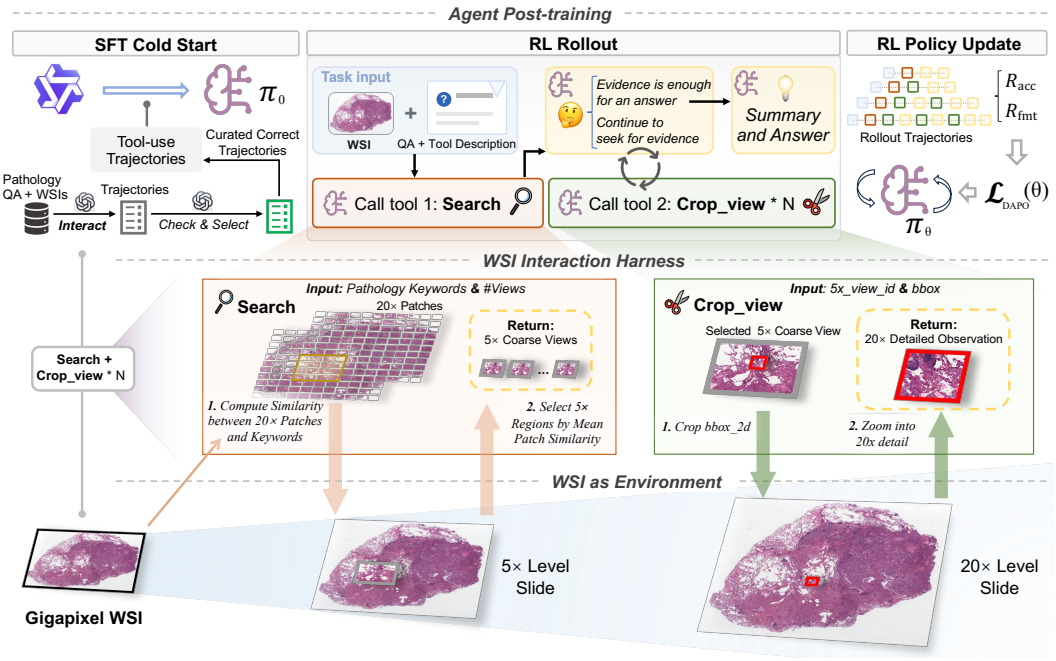

Figure 2: **Learning a WSI visual search and reasoning policy through environment interaction.** The policy interacts with the WSI through `Search` and `Crop_view`, implemented by the WSI harness. `Search` returns $5\times$ coarse views and `Crop_view` returns $20\times$ detailed observations from selected local areas. SFT initializes reasoning and tool calling from complete interaction trajectories. Online RL updates the policy using task outcomes from its own slide interactions.

3.3 SFT COLD START

SFT initializes reasoning followed by valid tool calls and continued interpretation of returned observations. We use complete tool-use trajectories collected through the same `Search`/`Crop_view` interface. A teacher MLLM (GPT-5.6-Sol-Medium) demonstrates WSI interaction by using `Search` to acquire observations and `Crop_view` to inspect local regions (Figure 2). Each trajectory records the reasoning, executed actions, returned observations, and terminal answer in their original order.

The sampled trajectories are curated before training. Deterministic checks verify trajectory executability, correct answers to closed questions, valid state transitions, and the pairing of each visual request with its returned observation. A text-only judge (GPT-5.6-Luna-Max) reviews semantic quality from the recorded trajectory. It also checks answers to open questions against the ground truth answers provided by the source datasets. We sample 55,000 tool-use trajectories from PathGen-Instruct and WSI-Bench, covering 5,000 WSIs. The final package contains 24,200 trajectories.

A retained interaction is serialized as one complete multimodal training sequence (Liu et al., 2023). Supervision covers reasoning before tool calling, `Search` and `Crop_view` requests, interpretation of returned observations, and the terminal answer. Tissue images and structured environment observations remain conditioning input. Each trajectory forms one training record and preserves the visual evidence available at every assistant turn. If $\mathcal{A}(\tau)$ denotes the spans generated by the assistant in trajectory τ, in the training pipeline of cold start, we optimize

$$\mathcal{L}_{\mathrm{SFT}}(\theta) = -\mathbb{E}_{\tau\sim\mathcal{D}} \sum_{m\in\mathcal{A}(\tau)} \sum_{u\in m} \log \pi_\theta(x_u \mid x_{<u}, I_{<u}), \tag{3}$$

where $I_{<u}$ contains the tissue observations acquired before token u. Actions and observations retain their original order. Data construction and curation statistics are reported in Appendix B.

3.4 AGENTIC RL THROUGH WSI INTERACTION

Starting from the SFT initialization, the policy generates complete rollouts in the WSI environment (Figure 2). Each rollout begins with a pathology question and a `Search` request specifying pathology keywords and the number of coarse views to return. The WSI harness executes the tool-call request and inserts the coarse views into the multimodal context.

The policy reasons over these views before requesting `Crop_view` on a selected region or terminating with an answer. Each crop returns morphology at $20\times$ from the corresponding WSI location. The detailed observation is appended to the visual history before the next model turn.

A rollout can inspect multiple locations in the same coarse view or move between coarse views. Each new observation can change the subsequent reasoning and inspection decision. The rollout ends when the policy emits the terminal answer. We represent this process as a trajectory interleaving model turns and environment observations. Let y_t denote the content generated by the model at step t, including reasoning and a visual request or terminal answer. The trajectory probability is

$$p_\theta(y_{1:T} \mid q, S) = \prod_{t=1}^{T} \pi_\theta(y_t \mid \mathcal{C}_\rho(q, y_{<t}, o_{\le t})), \qquad (s_{t+1}, o_{t+1}) = \mathcal{E}(S, s_t, a_t). \tag{4}$$

The environment supplies o_{t+1} after each executable action. It enters the context before the next policy decision. Optimization covers `Search` generation, local inspection based on coarse views, reasoning over returned observations, stopping, and final answer generation.

We launch rollouts on 5,000 closed-form question–answer pairs associated with 3,000 distinct WSIs, which are included in SFT datasets. Each trajectory τ receives the reward

$$R(\tau) = R_{\text{acc}}(\tau) + R_{\text{fmt}}(\tau), \tag{5}$$

where R_{acc} measures final answer correctness and R_{fmt} combines protocol compliance with penalties for invalid actions. For each question, eight sampled rollouts form one group. Their rewards are normalized within the group, and the resulting trajectory advantage is applied to assistant spans. We use clipped policy optimization in the style of Decoupled Clip and Dynamic sAmpling Policy Optimization (DAPO) (Yu et al., 2025). Implementation details are reported in Appendix C.

4 EXPERIMENTAL SETTING

Tasks and slide sources. We evaluate WholeSight on 2,956 questions covering tissue recognition, metastasis assessment, grading, diagnosis, and detailed pathology questions. Training uses TCGA slides paired with questions from PathGen-Instruct and WSI-Bench (Sun et al., 2025; Liang et al., 2025; Weinstein et al., 2013). We group the test data by slide origin. BRACS, CAMELYON16, CAMELYON17, the Clinical Proteomic Tumor Analysis Consortium (CPTAC) cohort, the Genotype-Tissue Expression (GTEx) project, and PANDA contribute 2,066 questions on slides from outside TCGA (OOD). Four visual question answering (VQA) collections on TCGA slides—MultiPathQA-TCGA, Expert-VQA, SlideBenchVQA, and WSI-VQA—contribute 890 questions (IND) (Brancati et al., 2022; Litjens et al., 2018; Bandi et al., 2019; Carithers et al., 2015; Bulten et al., 2022; Buckley et al., 2026; Chen et al., 2025b;a). Each dataset or benchmark component is reported separately. The MultiPathQA (Buckley et al., 2026) markers in the tables identify benchmark membership independently of this source grouping. Question counts and task definitions are given in Appendix D. All of the questions and WSI slides belonging to patients represented in training dataset are excluded from evaluation.

Compared models. SlideChat, WSI-LLaVA, HistoSelect, PathAgent, and PathNavigate use their original WSI inputs and inference procedures (Chen et al., 2025b; Liang et al., 2025; Huang et al., 2026; Chen et al., 2026a; Yang et al., 2026). Lingshu-7B, MedGemma-1.5-4B-it, and Patho-R1-7B receive fixed visual evidence. Qwen3.5 generates pathology keywords to retrieve five candidate regions. Each region is read at $20\times$ over the entire $5\times$ coarse view, and each medical MLLM answers in one pass from these images. Qwen2.5-VL-7B, Qwen3-VL-8B, and the original Qwen3.5-9B model (Qwen Team, 2026) use the same `Search`/`Crop_view` interface as WholeSight. Their backbones are comparably sized. More format adaptation details are described in Appendix D.1.

Training and interaction comparisons. We compare the original Qwen3.5-9B checkpoint, SFT at step 378, and RL checkpoints at steps 104, 192, and 272 using the same evaluation questions and WSI harness. WholeSight uses RL step 272 throughout the main comparison and interaction ablations. Default evaluation uses the same interaction protocol, episode bounds, and full visual history as training. The training comparison changes model parameters while keeping the evaluation interface fixed. The local evidence controls keep candidate regions fixed and change how tissue at $20\times$ is presented to the model, either by exposing each region in full at $20\times$, cropping predefined centers, or allowing the policy to select local observations. The other groups vary the maximum number of tool calls, candidate count, or number of recent images retained in context. Each comparison includes the default WholeSight setting and WSI harness configuration as its reference. Retrieval encoder comparisons are reported with the implementation details in Appendix A.

Scoring. Accuracy counts correctly answered questions over the full evaluation set, including failures to produce a valid answer. Overall accuracy weights every question equally. For Lingshu, MedGemma, and Patho-R1, answer recovery maps free text predictions to the available choices using only the generated text and choice descriptions. The recovered choices are then compared with the gold labels. Unrecoverable predictions count as incorrect.

5 Results and Analysis

5.1 Performance across slide sources and tasks

Table 1: **Performance across slide sources and tasks.** Accuracy (%).

System	OOD (WSI $\notin$ TCGA)						IND (WSI $\in$ TCGA)				Overall
	BRACS	C16	C17	CPTAC	GTEx†	PANDA†	TCGA†	Expert†	Slide†	WQ	
WSI systems of comparable scale											
SlideChat	13.53	40.10	11.80	51.67	8.20	12.69	0.00	38.28	<u>68.53</u>	50.87	27.60
WSI-LLaVA	27.61	41.10	57.00	40.83	27.32	20.81	14.93	46.09	62.94	50.58	39.88
HistoSelect	24.50	59.90	<u>63.60</u>	50.42	8.20	11.68	10.86	40.63	**70.05**	**60.47**	<u>43.03</u>
PathAgent	6.58	53.88	17.60	19.58	<u>48.63</u>	19.80	11.76	45.31	48.22	50.00	29.26
PathNavigate	9.87	60.40	47.20	39.58	28.96	15.74	0.00	42.97	60.91	48.55	35.59
Lingshu-7B	22.49	<u>72.18</u>	31.00	41.25	25.14	22.84	8.14	26.56	28.43	50.00	35.05
MedGemma-1.5-4B	17.55	65.41	62.00	38.33	17.49	27.41	1.81	30.47	30.96	45.64	37.42
Patho-R1-7B	<u>29.25</u>	64.41	58.20	47.92	10.93	<u>28.43</u>	1.81	28.13	33.50	46.51	39.41
Qwen2.5-VL-7B	14.44	60.15	57.60	40.83	8.74	20.81	7.69	29.69	31.47	23.84	32.51
Qwen3-VL-8B	23.58	49.37	47.00	<u>52.92</u>	15.30	13.20	2.71	<u>50.00</u>	46.19	37.50	34.91
Qwen3.5-9B	23.40	70.43	61.20	33.33	42.62	20.30	<u>20.81</u>	45.31	46.19	39.83	42.12
WholeSight (9B)	**39.85**	**79.70**	**65.80**	**53.33**	**57.38**	**30.46**	**35.29**	**56.25**	56.85	<u>52.33</u>	**54.13**

† Subsets of **MultiPathQA**. **C16/C17**: CAMELYON16/17; **Expert/Slide**: Expert-VQA/SlideBenchVQA; **WQ**: WSI-VQA choice subset. Bold and underlining mark the best and second best accuracies in each column; ties share a rank. Groups retain original WSI procedures, receive fixed regional images, or use the WholeSight visual interface, as Section 4 shows.

WholeSight achieves the overall best performance among the evaluated WSI systems of comparable scale (Table 1). Its 54.13% overall accuracy exceeds HistoSelect, the strongest baseline, by 11.10 points, WSI-LLaVA by 14.25 points, and the existing agents PathNavigate and PathAgent by 18.54 and 24.87 points. WholeSight ranks first in eight of ten evaluations, covering diagnosis, metastasis assessment, organ recognition, grading, and pathology questions. It also ranks second on WSI-VQA.

WholeSight improves over each comparison group. It reaches 39.85% on BRACS, 10.60 points above the next best model, and 79.70% on CAMELYON16, a margin of 7.52 points. On the TCGA component of MultiPathQA, its 35.29% exceeds the next best result by 14.48 points. Using the same `Search`/`Crop_view` interface, WholeSight exceeds Qwen2.5-VL-7B (32.51%) and Qwen3-VL-8B (34.91%) by 21.62 and 19.22 points. The medical MLLMs given fixed regional images reach 35.05% to 39.41% overall. HistoSelect gets the highest scores on SlideBenchVQA and WSI-VQA.

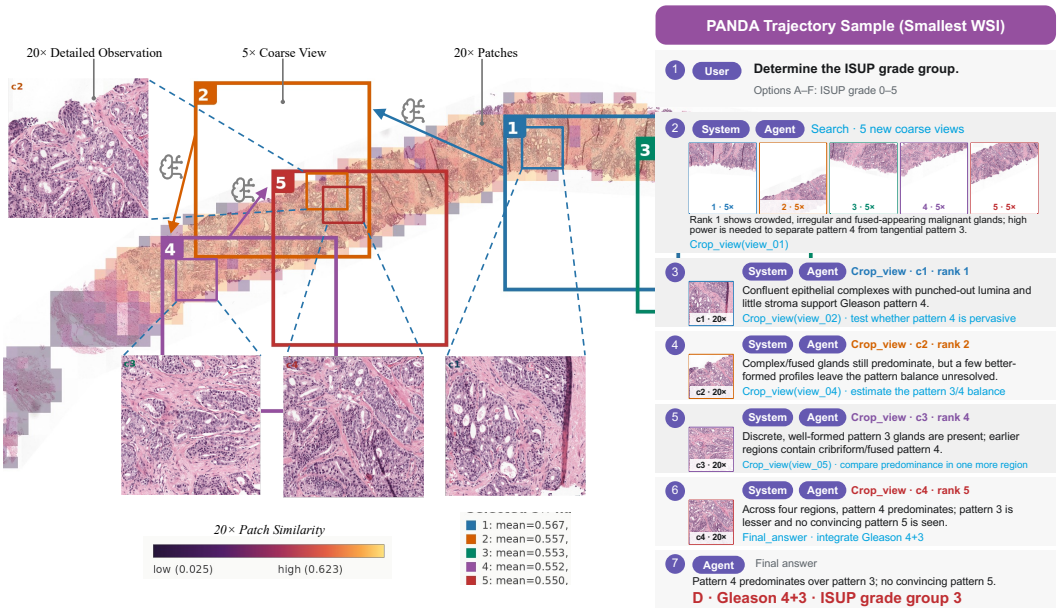

Figure 3: **Test-time interaction on a PANDA WSI.** The episode starts from the coarse regions returned by `Search` and proceeds through local inspections selected by the policy through `Crop_view` at 20×. The left panel marks the inspected tissue locations, and the right panel shows the recorded interactions with the WSI leading to the final International Society of Urological Pathology (ISUP) grade-group answer.

Generalization across slide sources. The same WholeSight weights and `Search`/`Crop_view` interface are used for every test collection. Relative to the original Qwen3.5-9B model with that interface, accuracy improves throughout the evaluation, from a gain of 4.60 points on CAMELYON17 to 20.00 points on CPTAC. On slides from outside TCGA, accuracy rises from 44.19% to 56.05%; on TCGA slides, it rises from 37.30% to 49.66%. These source groups gain 11.86 and 12.36 points, respectively. The improvement extends to organ recognition in GTEx and grading in PANDA, at 14.75 and 10.16 points over the original model. The same learned policy is used across all of these tasks and slide collections without changing its weights or interaction interface.

5.2 Test-time WSI interaction

Figure 3 shows interaction on a PANDA test WSI. The episode begins with the five coarse views returned by `Search`. Each local inspection is selected after the preceding observations have entered the model context, and the policy determines the next action by the current evidence state. The policy examines glandular morphology across several selected regions before producing the final ISUP grade-group answer. More interaction cases are shown in Appendix E.

Table 2: **Interaction training across evaluations.** Accuracy (%).

Training	OOD						IND			
	BRACS	C16	C17	CPTAC	GTEx†	PANDA†	TCGA†	Expert†	Slide†	WQ
Qwen3.5-9B	23.40	70.43	61.20	33.33	42.62	20.30	20.81	45.31	46.19	39.83
SFT	28.52	70.43	63.00	41.25	49.18	17.77	21.72	51.56	49.24	44.48
RL, 104 steps	35.47	81.70	41.80	47.92	48.09	26.90	14.03	53.13	54.31	49.71
RL, 192 steps	40.95	79.20	65.00	42.08	47.54	27.41	23.53	53.13	55.33	52.62
RL, 272 steps	**39.85**	**79.70**	**65.80**	**53.33**	**57.38**	**30.46**	**35.29**	**56.25**	**56.85**	**52.33**

† Subsets of **MultiPathQA**. **C16/C17**: CAMELYON16/17; **Expert/Slide**: Expert-VQA/SlideBenchVQA; **WQ**: WSI-VQA choice subset. Column abbreviations follow Table 1. RL checkpoints continue from SFT at step 378.

5.3 LEARNING THROUGH WSI INTERACTION

With the backbone, WSI harness, and test questions fixed, overall accuracy rises from 42.12% for the original model to 45.33% after SFT and 54.13% after online RL (Figure 4). Table 2 shows the corresponding results for all ten evaluations. SFT initializes executable interaction, but its task gains are uneven: CAMELYON16 remains at 70.43%, while PANDA changes from 20.30% to 17.77%. After online interaction training, these tasks reach 79.70% and 30.46%, and the final policy exceeds SFT in every evaluation. The largest gains over SFT are 13.57 points on MultiPathQA-TCGA, 12.08 on CPTAC, and 11.33 on BRACS.

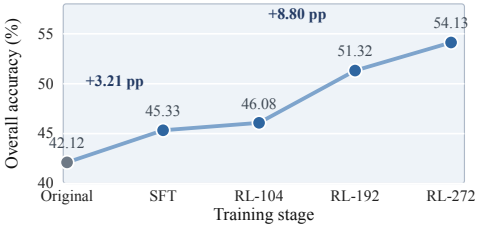

Figure 4: Accuracy during training, with SFT and RL gains marked.

The same model weights and observation rules are used across tissue recognition, grading, and diagnosis, and the gains extend across both slide-source groups.

Overall accuracy increases from 46.08% at 104 RL steps to 51.32% at 192 and 54.13% at 272. The task profiles do not improve in lockstep: CAMELYON16 is highest at 104 steps, while eight evaluations improve from step 192 to 272. Over that interval, CPTAC rises from 42.08% to 53.33%, MultiPathQA-TCGA from 23.53% to 35.29%, and PANDA from 27.41% to 30.46% (Table 2).

5.4 VISUAL EVIDENCE AND INTERACTION BUDGET

Selecting local evidence. With the trained model and initial retrieved regions fixed, this comparison separates high magnification content from the policy's choice of where to inspect. `Search` alone reaches 50.61% from the $5\times$ coarse views (Table 4). As Table 3 shows, supplying each candidate in full at $20\times$ gives 47.56%, and four crops at fixed centers give 50.07%. Adaptive inspection reaches 54.13%, gains of 6.57 and 4.06 points over these two fixed strategies. Simply presenting more high magnification tissue does not improve aggregate accuracy in these controls. The gain appears when the model chooses local tissue after reading the context including coarse views and can use each returned observation before deciding where to look next.

Adaptive inspection exceeds fixed centers in nine of the ten evaluations. BRACS rises from 35.10% to 39.85%, and MultiPathQA-TCGA rises from 14.03% to 35.29%. The whole candidate region input at $20\times$ is especially weak on BRACS at 17.18%. CAMELYON17 is the exception, where fixed centers reach 66.80% and adaptive inspection reaches 65.80%.

Table 3: **Local evidence acquisition.** Accuracy (%) with the trained WholeSight policy and retrieved coarse views of candidate regions held fixed.

Setting	OOD (WSI $\notin$ TCGA)						IND (WSI $\in$ TCGA)				Overall	Δ
	BRACS	C16	C17	CPTAC	GTEx†	PANDA†	TCGA†	Expert†	Slide†	WQ		
Full candidate regions at $20\times$	17.18	70.68	66.20	50.00	54.10	27.41	31.67	58.59	53.30	51.16	47.56	-6.57
Four crops at fixed centers	35.10	76.19	66.80	50.83	52.46	27.41	14.03	50.00	53.30	51.74	50.07	-4.06
Adaptive inspection (ours)	**39.85**	**79.70**	**65.80**	**53.33**	**57.38**	**30.46**	**35.29**	**56.25**	**56.85**	**52.33**	**54.13**	**0.00**

† Subsets of **MultiPathQA**. **C16/C17**: CAMELYON16/17; **Expert/Slide**: Expert-VQA/SlideBenchVQA; **WQ**: WSI-VQA choice subset. Δ is the Overall change relative to adaptive inspection. The fixed-center control uses four local crops.

Number of interactions. Three calls raise accuracy from 50.61% with `Search` alone to 52.98%, using at most two local inspections (Table 4). Six and eight calls reach 54.36% and 54.26%, close to the default 54.13%. WSI-VQA favors the default protocol, at 52.33% versus 46.80% and 47.67% under six and eight calls, while CAMELYON16 favors eight calls.

Coarse views. The candidate count changes the regional context with coarse views available before local inspection (Table 4). Increasing it from three to five raises overall accuracy from 51.96% to 54.13%. Eight views give 53.42%. GTEx rises from 46.45% to 57.38% between three and five views, and MultiPathQA-TCGA rises from 23.53% to 35.29%. Eight views help CAMELYON17 but reduce CPTAC and WSI-VQA relative to five views.

Visual history. Keeping only the latest four, six, or ten images yields 53.08%, 53.82%, and 54.33%, compared with 54.13% for full history (Table 4). These limits remove older coarse views and detailed observations from context as new images arrive. Truncating the image history costs at most 1.05 points, while removing subsequent local inspection costs 3.52 points. Full history gives the strongest WSI-VQA result among these variants at 52.33%.

Table 4: **Interaction budget and context.** Accuracy (%) with the trained WholeSight policy fixed. Configuration of tools and context management changes.

Setting	OOD (WSI $\notin$ TCGA)						IND (WSI $\in$ TCGA)				Overall	Δ
	BRACS	C16	C17	CPTAC	GTEx†	PANDA†	TCGA†	Expert†	Slide†	WQ		
Maximum tool calls												
1 (Search only)	29.25	72.18	66.60	52.92	58.47	24.87	33.03	58.59	55.33	50.87	50.61	-3.52
3 calls	37.48	81.70	65.80	52.50	54.10	28.43	32.13	57.81	55.33	49.71	52.98	-1.15
6 calls	42.41	80.70	66.60	52.08	56.83	32.99	36.20	57.03	56.85	46.80	54.36	+0.23
8 calls	40.40	85.21	66.20	52.92	55.19	31.47	33.48	53.91	58.38	47.67	54.26	+0.13
Default	**39.85**	**79.70**	**65.80**	**53.33**	**57.38**	**30.46**	**35.29**	**56.25**	**56.85**	**52.33**	**54.13**	**0.00**
Coarse views												
3 views	35.10	82.21	65.00	54.58	46.45	28.93	23.53	59.38	57.87	51.16	51.96	-2.17
5 views	**39.85**	**79.70**	**65.80**	**53.33**	**57.38**	**30.46**	**35.29**	**56.25**	**56.85**	**52.33**	**54.13**	**0.00**
8 views	39.12	74.94	69.00	50.42	56.83	28.43	35.75	58.59	57.36	50.29	53.42	-0.71
Visual history												
Latest 4 images	39.31	80.20	66.40	51.67	55.19	25.89	32.13	53.13	56.85	50.87	53.08	-1.05
Latest 6 images	40.04	79.45	67.40	51.67	57.92	29.44	34.84	58.59	56.35	48.55	53.82	-0.31
Latest 10 images	41.32	82.96	66.40	52.92	54.10	26.90	37.10	55.47	58.38	49.42	54.33	+0.20
Full history	**39.85**	**79.70**	**65.80**	**53.33**	**57.38**	**30.46**	**35.29**	**56.25**	**56.85**	**52.33**	**54.13**	**0.00**

† Subsets of **MultiPathQA**. **C16/C17**: CAMELYON16/17; **Expert/Slide**: Expert-VQA/SlideBenchVQA; **WQ**: WSI-VQA choice subset. Bold rows mark WholeSight's default settings; Δ is the Overall change from each default. Call limits include the initial `Search`. History limits retain the latest N returned images; full history matches training.

6 Conclusion

WholeSight treats gigapixel WSIs as interactive training environments for learning a general visual search and reasoning policy. The model learns to acquire tissue observations and interpret them as part of the same reasoning process. A WSI harness makes regional search and local inspection executable while keeping each high magnification observation connected to its low magnification tissue context. SFT initializes reasoning and tool calling from complete interaction trajectories, followed by online reinforcement learning through the policy's own slide interactions.

The resulting 9B model achieves 54.13% overall accuracy, exceeding the strongest evaluated system of comparable scale by 11.10 percentage points. It ranks first in eight evaluations and second on WSI-VQA. With the same weights and interaction interface, improvements over the original backbone extend across all tested tasks and slide collections, including sources outside TCGA. Online interaction training improves the SFT initialization in every evaluation. The visual access experiments also favor selecting local evidence as reasoning proceeds over fixed observation strategies.

WholeSight brings thinking with whole slide images into the learning process. Acquiring evidence and reasoning over it can be trained together through experience with WSIs, yielding a policy that transfers across the tasks and slide sources studied here.

AI USE STATEMENT

Generative AI assisted software implementation and review, dataset transformation, interpretation of experimental records, literature discovery, manuscript revision, and figure preparation. Teacher models (GPT-5.6) generated the supervised tool-use trajectories described in Section 3. The authors are responsible for the final manuscript, data handling, software, experimental interpretation, and disclosure of actual use.

ETHICS STATEMENT

The work develops methods for computational pathology research. It uses research WSI resources and their associated question and label collections. Data use is governed by the corresponding licenses, access conditions, and applicable privacy and institutional requirements. The study reports the scientific behavior of WSI reasoning agents, with model outputs interpreted in the context of the corresponding research tasks.

REPRODUCIBILITY STATEMENT

The supplementary material specifies data sources, trajectory construction, the WSI harness, optimization settings, and evaluation protocols. The result workbook and its normalized records preserve the reported accuracies and available integer counts. A table generation script computes aggregate scores where needed, derives the reported differences, and checks the manuscript tables against these records. Prompt templates and examples of complete interactions are included in the appendix.

REFERENCES

Peter Bandi, Oscar Geessink, Quirine Manson, Marcory Van Dijk, Maschenka Balkenhol, Meyke Hermsen, Babak Ehteshami Bejnordi, Byungjae Lee, Kyunghyun Paeng, Aoxiao Zhong, Quanzheng Li, Farhad Ghazvinian Zanjani, Svitlana Zinger, Keisuke Fukuta, Daisuke Komura, Vlado Ovtcharov, Shenghua Cheng, Shaoqun Zeng, Jeppe Thagaard, Anders B. Dahl, Huangjing Lin, Hao Chen, Ludwig Jacobsson, Martin Hedlund, Melih cetin, Eren Halici, Hunter Jackson, Richard Chen, Fabian Both, Jorg Franke, Heidi Kusters-Vandevelde, Willem Vreuls, Peter Bult, Bram van Ginneken, Jeroen van der Laak, and Geert Litjens. From detection of individual metastases to classification of lymph node status at the patient level: The camelyon17 challenge. *IEEE Transactions on Medical Imaging*, 38(2), 2019. ISSN 1558-254X. doi: 10.1109/tmi.2018.2867350. URL http://dx.doi.org/10.1109/TMI.2018.2867350.

Nadia Brancati, Anna Maria Anniciello, Pushpak Pati, Daniel Riccio, Giosuè Scognamiglio, Guillaume Jaume, Giuseppe De Pietro, Maurizio Di Bonito, Antonio Foncubierta, Gerardo Botti, Maria Gabrani, Florinda Feroce, and Maria Frucci. Bracs: A dataset for breast carcinoma subtyping in h&e histology images. *Database*, 2022, 2022. ISSN 1758-0463. doi: 10.1093/database/baac093. URL http://dx.doi.org/10.1093/database/baac093.

Thomas A. Buckley, Kian R. Weihrauch, Katherine Latham, Andrew Z. Zhou, Padmini A. Manrai, and Arjun K. Manrai. Navigating gigapixel pathology images with large multimodal models, 2026. URL https://arxiv.org/abs/2511.19652.

Wouter Bulten, Kimmo Kartasalo, Po-Hsuan Cameron Chen, Peter Ström, Hans Pinckaers, Kunal Nagpal, Yuannan Cai, David F Steiner, Hester Van Boven, Robert Vink, et al. Artificial intelligence for diagnosis and gleason grading of prostate cancer: the panda challenge. *Nature medicine*, 28(1): 154–163, 2022.

Latarsha J Carithers, Kristin Ardlie, Mary Barcus, Philip A Branton, Angela Britton, Stephen A Buia, Carolyn C Compton, David S DeLuca, Joanne Peter-Demchok, Ellen T Gelfand, et al. A novel approach to high-quality postmortem tissue procurement: the gtex project. *Biopreservation and biobanking*, 13(5):311–319, 2015.

Jingyun Chen, Linghan Cai, Zhikang Wang, Yi Huang, Songhan Jiang, Shenjin Huang, Hongpeng Wang, and Yongbing Zhang. Toward interpretable analysis of whole-slide pathology images via

large language model-based agentic reasoning. In *European Conference on Computer Vision*. Springer, 2026a.

Jingyun Chen, Fengchun Liu, Linghan Cai, Songhan Jiang, Shenjin Huang, Hongpeng Wang, Lequan Yu, and Yongbing Zhang. Agentic visual reasoning in whole-slide pathology images via active perception, 2026b. URL https://arxiv.org/abs/2608.08648.

Pingyi Chen, Chenglu Zhu, Sunyi Zheng, Honglin Li, and Lin Yang. Wsi-vqa: Interpreting whole slide images by generative visual question answering. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision – ECCV 2024*, pp. 401–417, Cham, 2025a. Springer Nature Switzerland. ISBN 978-3-031-72764-1.

Richard J. Chen, Chengkuan Chen, Yicong Li, Tiffany Y. Chen, Andrew D. Trister, Rahul G. Krishnan, and Faisal Mahmood. Scaling vision transformers to gigapixel images via hierarchical self-supervised learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16144–16155, June 2022.

Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3):850–862, 2024.

Ying Chen, Guoan Wang, Yuanfeng Ji, Yanjun Li, Jin Ye, Tianbin Li, Ming Hu, Rongshan Yu, Yu Qiao, and Junjun He. Slidechat: A large vision-language assistant for whole-slide pathology image understanding. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5134–5143, June 2025b.

Tong Ding, Sophia J. Wagner, Andrew H. Song, Richard J. Chen, Ming Y. Lu, Andrew Zhang, Anurag J. Vaidya, Guillaume Jaume, Muhammad Shaban, Ahrong Kim, Drew F. K. Williamson, Harry Robertson, Bowen Chen, Cristina Almagro-Pérez, Paul Doucet, Sharifa Sahai, Chengkuan Chen, Christina S. Chen, Daisuke Komura, Akihiro Kawabe, Mieko Ochi, Shinya Sato, Tomoyuki Yokose, Yohei Miyagi, Shumpei Ishikawa, Georg Gerber, Tingying Peng, Long Phi Le, and Faisal Mahmood. A multimodal whole-slide foundation model for pathology. *Nature Medicine*, 31 (11):3749–3761, Nov 2025. ISSN 1546-170X. doi: 10.1038/s41591-025-03982-3. URL https://doi.org/10.1038/s41591-025-03982-3.

Jiazhan Feng, Shijue Huang, Xingwei Qu, Ge Zhang, Yujia Qin, Baoquan Zhong, Chengquan Jiang, Jinxin Chi, and Wanjun Zhong. Retool: Reinforcement learning for strategic tool use in llms. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 37909–37926, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/4038c9208dfc22644c60ad39c24e5c53-Paper-Conference.pdf.

Jack Hong, Chenxiao Zhao, ChengLIn Zhu, Weiheng Lu, Guohai Xu, and XingYu. Deepeyesv2: Toward agentic multimodal model. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 114851–114872, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/badbef7d421640fb038399e9d7271b1c-Paper-Conference.pdf.

Yushi Hu, Weijia Shi, Xingyu Fu, Dan Roth, Mari Ostendorf, Luke Zettlemoyer, Noah Smith, and Ranjay Krishna. Visual sketchpad: Sketching as a visual chain of thought for multimodal language models. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 139348–139379. Curran Associates, Inc., 2024. doi: 10.52202/079017-4423. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/fb82011040977c7712409fbdb5456647-Paper-Conference.pdf.

Wentao Huang, Weimin Lyu, Peiliang Lou, Qingqiao Hu, Xiaoling Hu, Shahira Abousamra, Wenchao Han, Ruifeng Guo, Jiawei Zhou, Chao Chen, and Chen Wang. Act like a pathologist: Tissue-aware whole slide image reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6972–6981, June 2026.

Zhi Huang, Federico Bianchi, Mert Yuksekgonul, Thomas J. Montine, and James Zou. A visual–language foundation model for pathology image analysis using medical twitter. *Nature Medicine*, 29(9):2307–2316, Sep 2023. ISSN 1546-170X. doi: 10.1038/s41591-023-02504-3. URL https://doi.org/10.1038/s41591-023-02504-3.

Wisdom Ikezogwo, Saygin Seyfioglu, Fatemeh Ghezloo, Dylan Geva, Fatwir Sheikh Mohammed, Pavan Kumar Anand, Ranjay Krishna, and Linda Shapiro. Quilt-1m: One million image-text pairs for histopathology. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 37995–38017. Curran Associates, Inc., 2023. doi: 10.52202/075280-1654. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/775ec578876fa6812c062644964b9870-Paper-Datasets_and_Benchmarks.pdf.

Maximilian Ilse, Jakub Tomczak, and Max Welling. Attention-based deep multiple instance learning. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2127–2136. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/ilse18a.html.

Songhan Jiang, Fengchun Liu, Ziyue Wang, Linghan Cai, and Yongbing Zhang. Pathreasoner-r1: Instilling structured reasoning into pathology vision-language model via knowledge-guided policy optimization, 2026a. URL https://arxiv.org/abs/2601.21617.

Zheng Jiang, Heng Guo, Chengyu Fang, Changchen Xiao, Xinyang Hu, Lifeng Sun, and Minfeng Xu. Medvr: Annotation-free medical visual reasoning via agentic reinforcement learning. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 123025–123054, 2026b. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/c7f3c067c55ed888c1a5e1be77f3faa7-Paper-Conference.pdf.

Bowen Jin, Hansi Zeng, Zhenrui Yue, Jinsung Yoon, Sercan Arik, Dong Wang, Hamed Zamani, and Jiawei Han. Search-r1: Training llms to reason and leverage search engines with reinforcement learning, 2025. URL https://arxiv.org/abs/2503.09516.

Yuci Liang, Xinheng Lyu, Wenting Chen, Meidan Ding, Jipeng Zhang, Xiangjian He, Song Wu, Xiaohan Xing, Sen Yang, Xiyue Wang, and Linlin Shen. Wsi-llava: A multimodal large language model for whole slide image. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 22718–22727, October 2025.

Dankai Liao, Tianyi Zhang, Yufeng Wu, Xinyue Zhang, Qiaochu Xue, Zeyu Liu, Dachun Zhao, Linghan Cai, and Yueming Jin. Evipathbench: Benchmarking evidence acquisition and reasoning in vision-language models for whole-slide pathology, 2026. URL https://arxiv.org/abs/2607.19261.

Geert Litjens, Peter Bandi, Babak Ehteshami Bejnordi, Oscar Geessink, Maschenka Balkenhol, Peter Bult, Altuna Halilovic, Meyke Hermsen, Rob van de Loo, Rob Vogels, Quirine F Manson, Nikolas Stathonikos, Alexi Baidoshvili, Paul van Diest, Carla Wauters, Marcory van Dijk, and Jeroen van der Laak. 1399 h&e-stained sentinel lymph node sections of breast cancer patients: the camelyon dataset. *GigaScience*, 7(6), 2018. ISSN 2047-217X. doi: 10.1093/gigascience/giy065. URL http://dx.doi.org/10.1093/gigascience/giy065.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 34892–34916. Curran Associates, Inc., 2023. doi: 10.52202/075280-1516. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/6dcf277ea32ce3288914faf369fe6de0-Paper-Conference.pdf.

Ziyu Liu, Yuhang Zang, Yushan Zou, Zijian Liang, Xiaoyi Dong, Yuhang Cao, Haodong Duan, Dahua Lin, and Jiaqi Wang. Visual agentic reinforcement fine-tuning, 2025. URL https://arxiv.org/abs/2505.14246.

Meng Lu, Yuxing Lu, Yuchen Zhuang, Megan Mullins, Yang Xie, Guanghua Xiao, Charles Fleming, Wenqi Shi, and Xuan Wang. Medvistagym: A scalable training environment for thinking with medical images via tool-integrated reinforcement learning, 2026. URL https://arxiv.org/abs/2601.07107.

Ming Y. Lu, Drew F. K. Williamson, Tiffany Y. Chen, Richard J. Chen, Matteo Barbieri, and Faisal Mahmood. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nature Biomedical Engineering*, 5(6):555–570, Jun 2021. ISSN 2157-846X. doi: 10.1038/s41551-020-00682-w. URL https://doi.org/10.1038/s41551-020-00682-w.

Ming Y. Lu, Bowen Chen, Drew F. K. Williamson, Richard J. Chen, Ivy Liang, Tong Ding, Guillaume Jaume, Igor Odintsov, Long Phi Le, Georg Gerber, Anil V. Parwani, Andrew Zhang, and Faisal Mahmood. A visual-language foundation model for computational pathology. *Nature Medicine*, 30(3):863–874, Mar 2024a. ISSN 1546-170X. doi: 10.1038/s41591-024-02856-4. URL https://doi.org/10.1038/s41591-024-02856-4.

Ming Y. Lu, Bowen Chen, Drew F. K. Williamson, Richard J. Chen, Melissa Zhao, Aaron K. Chow, Kenji Ikemura, Ahrong Kim, Dimitra Pouli, Ankush Patel, Amr Soliman, Chengkuan Chen, Tong Ding, Judy J. Wang, Georg Gerber, Ivy Liang, Long Phi Le, Anil V. Parwani, Luca L. Weishaupt, and Faisal Mahmood. A multimodal generative ai copilot for human pathology. *Nature*, 634 (8033):466–473, Oct 2024b. ISSN 1476-4687. doi: 10.1038/s41586-024-07618-3. URL https://doi.org/10.1038/s41586-024-07618-3.

Chi Phan, Tianyi Zhang, Yufeng Wu, Qiaochu Xue, Jiajie Zhang, Linghan Cai, Zeyu Liu, Sudong Wang, Yueming Jin, and Dan Hu. Pathscale-r1: Cross-scale reasoning for pathological image analysis, 2026. URL https://arxiv.org/abs/2607.23794.

Qwen Team. Qwen3.5: Towards native multimodal agents, February 2026. URL https://qwen.ai/blog?id=qwen3.5.

Timo Schick, Jane Dwivedi-Yu, Roberto Dessi, Roberta Raileanu, Maria Lomeli, Eric Hambro, Luke Zettlemoyer, Nicola Cancedda, and Thomas Scialom. Toolformer: Language models can teach themselves to use tools. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 68539–68551. Curran Associates, Inc., 2023. doi: 10.52202/075280-2997. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/d842425e4bf79ba039352da0f658a906-Paper-Conference.pdf.

Mehmet Saygin Seyfioglu, Wisdom O. Ikezogwo, Fatemeh Ghezloo, Ranjay Krishna, and Linda Shapiro. Quilt-llava: Visual instruction tuning by extracting localized narratives from open-source histopathology videos. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13183–13192, June 2024.

Zhuchen Shao, Hao Bian, Yang Chen, Yifeng Wang, Jian Zhang, Xiangyang Ji, and yongbing zhang. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 2136–2147. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/10c272d06794d3e5785d5e7c5356e9ff-Paper.pdf.

Alex Su, Haozhe Wang, Weiming Ren, Fangzhen Lin, and Wenhu Chen. Pixel reasoner: Incentivizing pixel space reasoning via curiosity-driven reinforcement learning. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 8222–8251. Curran Associates, Inc., 2025a. doi: 10.52202/085713-0280. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/0c38f54740062529aa4117a04b583f3c-Paper-Conference.pdf.

Zhaochen Su, Linjie Li, Mingyang Song, Yunzhuo Hao, Zhengyuan Yang, Jun Zhang, Guanjie Chen, Jiawei Gu, Juntao Li, Xiaoye Qu, and Yu Cheng. Openthinkimg: Learning to think with images via visual tool reinforcement learning, 2025b. URL https://arxiv.org/abs/2505.08617.

YUXUAN SUN, Yixuan Si, Chenglu Zhu, Kai Zhang, Zhongyi Shui, Bowen Ding, Tao Lin, and Lin Yang. Cpathagent: An agent-based foundation model for interpretable high-resolution pathology image analysis mimicking pathologists'diagnostic logic. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 101673–101731. Curran Associates, Inc., 2025. doi: 10.52202/085713-3404. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/933b5d002cf251b3e854d586e55ac58c-Paper-Conference.pdf.

Yuxuan Sun, Yunlong Zhang, Yixuan Si, Chenglu Zhu, Kai Zhang, Zhongyi Shui, Jingxiong Li, Xuan Gong, XINHENG LYU, Tao Lin, and Lin Yang. Pathgen-1.6m: 1.6 million pathology image-text pairs generation through multi-agent collaboration. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 94611–94653, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/ebf8764ecf0688cdd9fe1e5a9c525d0d-Paper-Conference.pdf.

Eugene Vorontsov, Alican Bozkurt, Adam Casson, George Shaikovski, Michal Zelechowski, Kristen Severson, Eric Zimmermann, James Hall, Neil Tenenholtz, Nicolo Fusi, Ellen Yang, Philippe Mathieu, Alexander van Eck, Donghun Lee, Julian Viret, Eric Robert, Yi Kan Wang, Jeremy D. Kunz, Matthew C. H. Lee, Jan H. Bernhard, Ran A. Godrich, Gerard Oakley, Ewan Millar, Matthew Hanna, Hannah Wen, Juan A. Retamero, William A. Moye, Razik Yousfi, Christopher Kanan, David S. Klimstra, Brandon Rothrock, Siqi Liu, and Thomas J. Fuchs. A foundation model for clinical-grade computational pathology and rare cancers detection. *Nature Medicine*, 30(10):2924–2935, Oct 2024. ISSN 1546-170X. doi: 10.1038/s41591-024-03141-0. URL https://doi.org/10.1038/s41591-024-03141-0.

Sheng Wang, Ruiming Wu, Charles Herndon, Songhao Li, Yihang Liu, Shunsuke Koga, Xiaowei Xu, David E. Elder, Jonathan Alex Miles, Annie Jin, Ikuko Hirai, Meaghan Dougher, Jeanne Shen, and Zhi Huang. Pathology-cot: learning visual chain-of-thought agents from expert whole-slide image diagnosis behaviour. *Nature Biomedical Engineering*, Jul 2026. ISSN 2157-846X. doi: 10.1038/s41551-026-01739-y. URL https://doi.org/10.1038/s41551-026-01739-y.

John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, and Joshua M Stuart. The cancer genome atlas pan-cancer analysis project. *Nature genetics*, 45(10):1113–1120, 2013.

Luca L. Weishaupt, Chengkuan Chen, Drew F. K. Williamson, Richard J. Chen, Guillaume Jaume, Tong Ding, Bowen Chen, Anurag Vaidya, Long Phi Le, Guillaume Jaume, Ming Y. Lu, and Faisal Mahmood. Evidence-based diagnostic reasoning with multi-agent copilot for human pathology, 2026. URL https://arxiv.org/abs/2506.20964.

Bryan Wong, Xun Xu, Huazhu Fu, Nancy F. Chen, and Mun Yong Yi. Beyond relevance: Bayesian evidence acquisition for agentic whole-slide image reasoning, 2026. URL https://arxiv.org/abs/2608.05757.

Penghao Wu and Saining Xie. V*: Guided visual search as a core mechanism in multimodal llms, 2023. URL https://arxiv.org/abs/2312.14135.

Hanwen Xu, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong, Zelalem Gero, Javier González, Yu Gu, Yanbo Xu, Mu Wei, Wenhui Wang, Shuming Ma, Furu Wei, Jianwei Yang, Chunyuan Li, Jianfeng Gao, Jaylen Rosemon, Tucker Bower, Soohee Lee, Roshanthi Weerasinghe, Bill J. Wright, Ari Robicsek, Brian Piening, Carlo Bifulco, Sheng Wang, and Hoifung Poon. A whole-slide foundation model for digital pathology from real-world data. *Nature*, 630(8015):181–188, Jun 2024. ISSN 1476-4687. doi: 10.1038/s41586-024-07441-w. URL https://doi.org/10.1038/s41586-024-07441-w.

Chunze Yang, Qidong Liu, Wenjie Zhao, Yue Tang, Jiusong Ge, Di Zhang, Jiashuai Liu, Lei Wu, Junbo Lu, Ni Zhang, Xian Wu, Zeyu Gao, and Chen Li. Pathnavigate: A training-free pathology agent with surprise-guided scan and shared slide memory for whole-slide image vqa, 2026. URL https://arxiv.org/abs/2605.23559.

Shunyu Yao, Jeffrey Zhao, Dian Yu, Nan Du, Izhak Shafran, Karthik Narasimhan, and Yuan Cao. React: Synergizing reasoning and acting in language models, 2023. URL https://arxiv.org/abs/2210.03629.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, juncai liu, LingJun Liu, Xin Liu, Haibin Lin, Zhiqi Lin, Bole Ma, Guangming Sheng, Yuxuan Tong, Chi Zhang, Mofan Zhang, Ru Zhang, Wang Zhang, Hang Zhu, Jinhua Zhu, Jiaze Chen, Jiangjie Chen, Chengyi Wang, Hongli Yu, Yuxuan Song, Xiangpeng Wei, Hao Zhou, Jingjing Liu, Wei-Ying Ma, Ya-Qin Zhang, Lin Yan, Yonghui Wu, and Mingxuan Wang. Dapo: An open-source llm reinforcement learning system at scale. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen (eds.), *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 113222–113244. Curran Associates, Inc., 2025. doi: 10.52202/085713-3775. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/a4277440d50f1f15d2cb4c14f7e0c0d2-Paper-Conference.pdf.

Wenchuan Zhang, Penghao Zhang, Jingru Guo, Tao Cheng, Jie Chen, Shuwan Zhang, Zhang Zhang, Yuhao Yi, and Hong Bu. Patho-r1: A multimodal reinforcement learning-based pathology expert reasoner. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(33), 2026. ISSN 2159-5399. doi: 10.1609/aaai.v40i33.40071. URL http://dx.doi.org/10.1609/aaai.v40i33.40071.

Ziwei Zheng, Minghao Yang, Jack Hong, Chenxiao Zhao, Guohai Xu, Le Yang, Chao Shen, and XingYu. Deepeyes: Incentivizing "thinking with images" via reinforcement learning. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 126775–126798, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/cdb347d7516d52a9280ea9d6708911f6-Paper-Conference.pdf.

Xiao Zhou, Luoyi Sun, Dexuan He, Wenbin Guan, Ge Wang, Ruifen Wang, Lifeng Wang, Xiaojun Yuan, Xin Sun, Ya Zhang, et al. Knowledge-enhanced pretraining for vision-language pathology foundation model on cancer diagnosis. *Cancer Cell*, 44(4):777–791, Apr 2026. ISSN 1535-6108. doi: 10.1016/j.ccell.2026.01.019. URL http://dx.doi.org/10.1016/j.ccell.2026.01.019.

A WSI Environment and Visual Access

Precomputed tissue patch features. Before interaction, we extract nonoverlapping 256×256 tissue patches at $20\times$ on a regular grid. KEEP encodes each patch into a feature vector with 768 dimensions (Zhou et al., 2026). We store each vector with the patch coordinates and dimensions in the original slide image, denoted level 0, along with the slide dimensions. The coordinates identify neighboring patches even where background regions leave gaps in the grid. KEEP remains fixed during policy training. Patch features are computed once and reused for `Search` on that slide.

From pathology keywords to patch scores. `Search` receives pathology keywords and the requested number of coarse views. Each returned coarse view depicts one selected candidate region. The KEEP text encoder embeds each keyword separately. Each embedding is normalized to unit length; their mean is then normalized again to obtain the retrieval query. Let $\mathbf{t}_j$ be the normalized embedding of keyword j and $\mathbf{f}_i$ the normalized feature of tissue patch i. For m keywords, the query and patch scores are

$$\mathbf{z} = \frac{\sum_{j=1}^{m} \mathbf{t}_j}{\left\| \sum_{j=1}^{m} \mathbf{t}_j \right\|_2}, \qquad s_i = \mathbf{f}_i^{\mathsf{T}} \mathbf{z}. \tag{6}$$

The keyword and patch embeddings share the KEEP feature space. We compute their cosine similarities in batches and place each score at its patch position on the grid. The resulting map records keyword similarity across the slide.

Retrieval encoder choice. KEEP provides the text and image embedding space used by the main WholeSight experiments. We replace it with PLIP or CONCH at evaluation time while keeping the WholeSight policy and WSI harness protocol fixed. This changes the similarity scores used to rank candidate regions before returning coarse views. KEEP yields 54.13% overall accuracy, compared with 51.29% for PLIP and 50.00% for CONCH (Table 5).

KEEP gives the highest overall accuracy, leading PLIP by 2.84 points and CONCH by 4.13 points. Its advantage is largest on BRACS, at 39.85% compared with 25.41% and 27.24%. It also gives higher accuracy on PANDA and WSI-VQA than either alternative. Some tasks favor another encoder, but KEEP provides the strongest aggregate performance for the trained policy. The training comparison in Table 2 holds this encoder fixed throughout.

Table 5: **Retrieval encoder ablation.** Accuracy (%) with the WholeSight policy and interaction protocol fixed.

Encoder	OOD (WSI $\notin$ TCGA)						IND (WSI $\in$ TCGA)				Overall	Δ
	BRACS	C16	C17	CPTAC	GTEx†	PANDA†	TCGA†	Expert†	Slide†	WQ		
KEEP (ours)	**39.85**	**79.70**	**65.80**	**53.33**	**57.38**	**30.46**	**35.29**	**56.25**	**56.85**	**52.33**	**54.13**	**0.00**
PLIP	25.41	79.20	63.60	54.58	58.47	26.90	35.29	65.63	58.38	50.87	51.29	-2.84
CONCH	27.24	77.94	64.20	47.92	55.19	27.41	30.32	54.69	61.42	49.13	50.00	-4.13

† Subsets of **MultiPathQA**. **C16/C17**: CAMELYON16/17; **Expert/Slide**: Expert-VQA/SlideBenchVQA; **WQ**: WSI-VQA choice subset. Δ is the change in Overall accuracy relative to KEEP, in percentage points. All encoders are evaluated on the same 2,956 questions.

From patch scores to candidate regions. The WSI harness evaluates every 16×16 window on the patch grid, moving by one patch between neighboring windows. Each window covers a candidate region of 4096×4096 pixels at $20\times$. The same tissue area occupies 1024×1024 pixels at $5\times$, or 8192×8192 level 0 pixels on a slide scanned at $40\times$. The conversion uses the stored patch coordinates and dimensions.

A candidate must contain tissue patches with stored features in at least 10% of its 256 grid positions, or at least 26 patches. Its score is the mean similarity over those tissue patches,

$$R(W) = \frac{1}{n_W} \sum_{i \in \mathcal{I}(W)} s_i, \qquad n_W = |\mathcal{I}(W)| \geq 26, \tag{7}$$

where $\mathcal{I}(W)$ contains the tissue patches with stored features in window W. Grid positions without a tissue feature do not enter the denominator. Summed area tables compute the similarity sum and patch count for the overlapping windows. Windows are ranked by $R(W)$. Ties favor windows with more tissue patches, then more patches whose scores are at or above the slide's 95th percentile. Remaining ties are resolved by ascending vertical and horizontal position.

Greedy nonmaximum suppression limits overlap among selected regions. A window is skipped when its intersection over union with an already selected window exceeds 0.25. Selection stops at the requested candidate count, which is five in the default setting. The returned set can be smaller when fewer windows satisfy these conditions.

Rendering the coarse views. For each selected window, the WSI harness reads the corresponding tissue from the slide pyramid and renders a 1024×1024 RGB image at $5\times$. The pyramid stores the slide at multiple sampling resolutions. OpenSlide selects the coarsest available level whose sampling is at least as fine as the requested output. The image is resampled to the output grid, with white padding for any area outside the slide. Each coarse view depicts one continuous candidate region. The policy receives these views together with their ranks and identifiers. The map of patch similarity scores is used only by the WSI harness.

Each view stores its level 0 bounding box and the coordinate transform between the rendered image and the slide. For a view covering $[X_0, Y_0, X_1, Y_1)$, a point (u, v) in its 1024×1024 image maps to

$$X = X_0 + \frac{u}{1024}(X_1 - X_0), \qquad Y = Y_0 + \frac{v}{1024}(Y_1 - Y_0). \tag{8}$$

The same transform maps the corners of a local crop request. Coordinates refer to pixel edges, with the upper bounds excluded.

Reading a detailed observation. `Crop_view` takes a coarse view identifier and a bounding box in its 1024×1024 image at $5\times$. Each box side is constrained to 256 to 1024 pixels in this $5\times$ coordinate system. A request outside the range is expanded or contracted around its center. The backend maps the adjusted box to level 0 and reads the tissue at $20\times$, multiplying each side length by four. The returned image spans 1024 to 4096 pixels per side at $20\times$. The effective box is shifted inside the slide bounds when necessary. The observation records the requested box, effective box, parent view, and any size or position adjustment.

Episode state and visual history. Each episode starts with one `Search` action. The policy can answer from the coarse views or request one or more detailed observations with `Crop_view`. After every crop, it may inspect another location in the same coarse view, move to another coarse view, or answer. No crop is required when the coarse views already provide enough evidence.

Default training and evaluation retain the full image sequence, beginning with the `Search` views and appending each successful crop. In the visual history ablation, the model context contains only the latest N images from this sequence, with N set to 4, 6, or 10. This limit applies to all returned images, so older coarse views and crops leave the visual context as new observations arrive. Table 4 keeps the model and interaction rules fixed while changing this image limit. Message formats and tool arguments are described in Appendix F.

Policy and harness boundary. WholeSight learns the MLLM interaction policy that generates `Search` requests, selects local inspections, interprets returned observations, and decides when to answer. Patch encoding, region ranking, rendering, and coordinate conversion are fixed operations of the WSI harness. Joint visual search and reasoning refers to reasoning and executable visual actions being produced by the same learned policy from the same evolving interaction history.

B Training Data and Trajectory Construction

SFT trajectory collection and filtering. Training questions come from PathGen-Instruct and WSI-Bench, whose associated WSIs are derived from TCGA (Sun et al., 2025; Liang et al., 2025; Weinstein et al., 2013). SFT uses complete interaction trajectories to initialize reasoning followed by valid tool calls and continued interpretation of the returned observations. The teacher uses the

same `Search`/`Crop_view` interface as WholeSight. It requests coarse views, may acquire detailed observations, and ends with a final answer. We sample 55,000 interaction trajectories covering 5,000 WSIs.

We check each trajectory for executable tool calls, valid state transitions, and correct pairing between requests and returned observations. Closed questions are also checked against their reference labels. A judge that reads only text reviews the recorded trajectory for semantic quality and evaluates open answers against the references supplied by the source datasets. The judge reads the recorded trajectory and does not acquire new WSI observations. After filtering, 24,200 complete trajectories are used for SFT.

Role of the SFT stage. The retained trajectories provide an interaction cold start for the student policy. Supervision follows the same reasoning, action, observation, and answer order used during rollout, including continued reasoning after tool returns. The teacher constructs the training trajectories and is not part of the test-time WSI agent.

Multimodal sequence construction. Each selected trajectory becomes one multimodal training sequence. Actions, returned tissue observations, intermediate reasoning, and the terminal answer keep their original order. At every assistant turn, the sequence contains the visual evidence acquired up to that point.

The SFT loss covers reasoning, `Search`/`Crop_view` requests, interpretation of returned observations, and the terminal answer. System and user tokens, tool observations, and image content are used as conditioning input. Padding is excluded from the loss. Appendix F shows how these messages are rendered for the model.

C Online Rollout and Training Configuration

Online training begins from the SFT initialization. The rollout set contains 5,000 closed QA pairs associated with 3,000 distinct WSIs. For each question, the policy samples eight rollouts. Model generation alternates with tool execution, and each returned observation is inserted before the next model turn.

Reward and relative advantage. Each trajectory receives the reward in Equation 5. R_{acc} is binary final answer correctness. R_{fmt} combines a bounded tool protocol score with penalties. The protocol score evaluates successful tool calls, argument validity, schema validity, and final answer validity. Process penalties apply to malformed, invalid, or duplicate actions.

For the eight rollouts generated from one question, let R_i denote the trajectory rewards. We compute

$$A_i = \frac{R_i - \bar{R}}{s_R + \epsilon}, \qquad \bar{R} = \frac{1}{8}\sum_{j=1}^{8} R_j, \tag{9}$$

where s_R is the sample standard deviation within the group. A group with no reward variance contributes zero relative advantage. The same A_i is assigned to the assistant spans generated by the model in trajectory i, including reasoning, tool requests, and the terminal answer. System and user messages, tool observations, and returned images remain conditioning context. We use asymmetric clipped policy optimization based on DAPO (Yu et al., 2025).

Role of policy optimization. Online RL is the optimization stage used after the SFT cold start. The answer and protocol rewards, group-relative advantage, and DAPO-based clipping are implementation choices for training the WSI interaction policy. The paper makes no claim of a new reward, RL objective, or policy optimization algorithm.

The training-stage comparison is reported in Table 2 and Figure 4. The original model, SFT initialization, and RL checkpoints use the same questions and WSI harness. The trained WholeSight policy is held fixed for the local evidence, interaction budget, candidate count, and visual history comparisons in Tables 3 and 4.

Table 6: **Online training configuration.**

Setting	Value
Rollouts per question	8
Training batch	18 prompts
Actor learning rate	10^{-6}, constant
Weight decay	0.01
Policy epochs per update	1
Loss reduction	Mean over optimized assistant tokens
Lower / upper ratio clipping	0.20 / 0.28
Dual clip parameter	10
KL reward / actor KL loss	Disabled / disabled
Entropy coefficient	0
Critic	Disabled
Vision tower	Trainable
Maximum response length	32,768 tokens
Maximum assistant turns	12
Parallel tool calls	1
Visual history	All returned images are retained

D Evaluation Suite and Comparison Protocols

WholeSight is evaluated as a learned MLLM interaction policy operating through a fixed WSI harness. The policy generates pathology queries, interprets the returned coarse views, selects local high-magnification observations, updates its reasoning after each observation, decides whether to continue, and produces the final answer, while patch encoding, patch scoring, candidate ranking, rendering, and coordinate conversion remain fixed. Accordingly, end-to-end learning refers to the policy loop and does not imply that every harness operation is jointly optimized. SFT provides the executable interaction cold start; the teacher is absent at test time, and its collection prompt does not contain the reference answer, which is introduced only after generation for checking and filtering. Online RL then continues training the same policy under the same interface. RL-272 is the terminal checkpoint of that completed run, and RL-104 and RL-192 are intermediate snapshots rather than checkpoints selected by evaluation performance. The 8.80-point SFT-to-RL improvement is therefore reported as a policy-level training outcome, not as a causal attribution to Crop, Search, the vision tower, the reward, or another individual component. Reward-component ablations, frozen-module controls, compute-matched SFT continuation, additional seeds, or human review of teacher trajectories would refine optimization attribution, training variance, or synthetic-data quality; these are separate questions from whether a single WSI interaction policy can be trained to couple evidence acquisition and reasoning.

The evaluation is organized around the same distinction. Qwen2.5-VL, Qwen3-VL, and the original Qwen3.5 use the same `Search`/`Crop_view` interface as WholeSight and provide the matched-interface controls for policy learning; methods using their own WSI procedures provide system-level context, and no direct performance claim is made against a method that is not included in the reported comparison table. The local-evidence experiments hold the trained model and retrieved candidates fixed and change only visual access after `Search`, so they test whether policy-selected local observations improve downstream answering. They do not claim lesion-localization accuracy, pathologist ROI agreement, oracle retrieval performance, or superiority over every possible inspection heuristic; random, saliency-based, uncertainty-based, information-gain, and oracle crop policies would instead compare alternative inspection heuristics. All evaluation questions and WSI slides from patients represented in training are excluded from evaluation. OOD and IND denote slide origin outside TCGA and within TCGA only, and Overall accuracy is the question-weighted micro-average over all 2,956 questions. The tool-call study measures sensitivity of task accuracy to the available number of actions; it is not an optimized latency or accuracy-cost study, and the paper does not make prospective clinical-utility or expert-level ROI-localization claims.

Evaluation suite. The evaluation suite contains 2,956 questions. Evaluation sources using slides from outside TCGA account for 2,066 questions, and sources using TCGA slides account for 890

questions. Table 7 lists the evaluation sources and their question counts, following the OOD/IND grouping and MultiPathQA markers in the main results.

Table 7: **Evaluation sources and question counts.**

Evaluation source	Domain	Questions	Task / question source
BRACS	OOD	547	Diagnosis among seven categories (Brancati et al., 2022)
CAMELYON16	OOD	399	Metastasis presence (Litjens et al., 2018)
CAMELYON17	OOD	500	Slide-level metastasis category (Bandi et al., 2019)
CPTAC	OOD	240	SlideBench-CPTAC questions (Chen et al., 2025b)
GTEx†	OOD	183	Organ recognition (Carithers et al., 2015; Buckley et al., 2026)
PANDA†	OOD	197	ISUP/Gleason grading (Bulten et al., 2022; Buckley et al., 2026)
MultiPathQA-TCGA†	IND	221	TCGA component of MultiPathQA (Buckley et al., 2026)
Expert-VQA†	IND	128	Questions written by experts (Buckley et al., 2026)
SlideBenchVQA†	IND	197	SlideBench questions in MultiPathQA (Chen et al., 2025b; Buckley et al., 2026)
WSI-VQA	IND	344	Choice questions (Chen et al., 2025a)
Total		2,956	

† Subsets of MultiPathQA. OOD and IND refer to slide source, following Section 4. **All of the questions and WSI slides that appear in training are excluded.**

Scope of the source grouping. OOD and IND describe slide origin only: outside TCGA versus TCGA. Evaluation still excludes every question and WSI used in training, including the TCGA-based evaluation components. The analysis therefore measures transfer across the reported slide collections and pathology tasks; the labels are not a broader semantic OOD taxonomy.

D.1 Baseline Inputs and Scoring

SlideChat, WSI-LLaVA, HistoSelect, PathAgent, and PathNavigate use their original WSI inputs and inference procedures (Chen et al., 2025b; Liang et al., 2025; Huang et al., 2026; Chen et al., 2026a; Yang et al., 2026). Table 8 summarizes the visual input used for each comparison group.

Table 8: **Visual input for the compared systems.**

System	Visual input and inference
SlideChat; WSI-LLaVA	Original WSI representation and slide-level inference
HistoSelect	Original tissue selection followed by SlideChat
PathAgent; PathNavigate	Original WSI agent procedure
Lingshu-7B; MedGemma-1.5-4B-it; Patho-R1-7B	Five complete candidate regions read at 20×; one answer without further `Crop_view`
Qwen2.5-VL-7B; Qwen3-VL-8B; original Qwen3.5-9B	The `Search`/`Crop_view` interface used by WholeSight

Coupling of visual access and reasoning. CPathAgent uses staged supervised training within a unified pathology model, while AdaptivePath trains a Navigator separately from downstream interpretation and deliberation (SUN et al., 2025; Chen et al., 2026b). PathAgent and PathNavigate organize iterative visual access through modular or inference-time procedures (Chen et al., 2026a; Yang et al., 2026). WholeSight uses one MLLM policy for reasoning, executable visual actions, follow-up inspection, stopping, and the final answer. The comparison concerns how evidence acquisition and reasoning are coupled during learning; retrieval and rendering remain fixed WSI harness operations.

For Lingshu, MedGemma, and Patho-R1, Qwen3.5 generates the pathology keywords used to retrieve five candidate regions. Each region is read from the WSI at $20\times$ over the same tissue area shown in its $5\times$ coarse view. The model receives all five images before answering. In adaptive inspection, the policy reviews the coarse views before choosing local regions.

For each evaluation source, accuracy uses all questions as its denominator. For Lingshu, MedGemma, and Patho-R1, answer recovery maps the generated text to the available choices using only that text and the choice descriptions. The recovered choice is then compared with the gold label. Unrecoverable predictions count as incorrect. The other systems use the benchmark answer parser. Overall accuracy is the number of correct answers divided by the full set of 2,956 questions.

Controlled interaction settings. The interaction ablations keep the trained WholeSight policy fixed. The local evidence comparison changes how tissue at high magnification is presented after `Search`. The full region control reads the entire tissue area of each candidate at $20\times$. The fixed center control crops the first four candidate regions using $[256, 256, 768, 768]$ in the $5\times$ coarse view. This box has side length 512 at $5\times$ and is read as 2048×2048 pixels at $20\times$, following Appendix A. The call budget comparison limits the total number of tool calls, including the initial `Search`. A limit of one requires an answer from the coarse views. Default evaluation matches the training protocol and episode bounds in Table 6; it imposes no additional evaluation cap. The policy can answer before reaching a limit. Once the imposed call limit is reached, the next turn is restricted to a terminal answer. The remaining comparisons change the candidate count or retain only the latest N images in context. Full history keeps all returned images, as in training.

Reading the interaction controls. The local evidence controls hold the trained policy and retrieved candidate set fixed while changing visual access after `Search`. Search only removes later local inspection; the full region and fixed center settings provide high magnification tissue without policy-selected locations; adaptive inspection lets the policy choose each local observation from its current interaction state. These comparisons isolate visual access behavior under the same trained policy, independent of changes to the reward or optimizer.

E Interaction Examples

Figure 3 presents one test-time PANDA interaction in the main paper. Here we show five additional visual summaries of WholeSight interactions on test WSIs from the evaluation suite of 2,956 questions. The plates link $5\times$ coarse views to $20\times$ detailed observations and retain the recorded final answers. Their annotations summarize the observations and actions. The CAMELYON16 plate presents selected observations from a longer episode; the other four plates include every crop observation in their respective episodes.

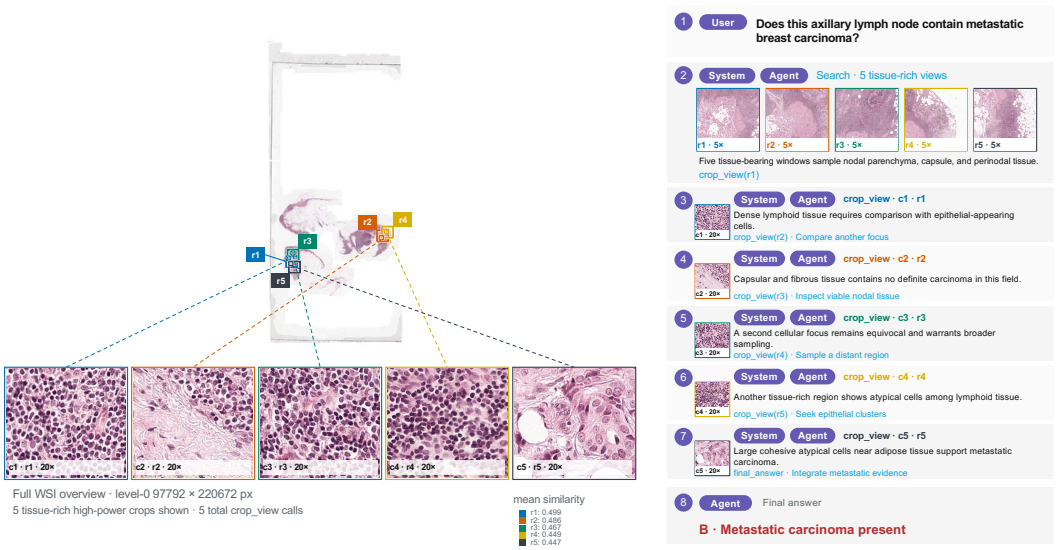

Figure 5: **Sampling nodal and perinodal tissue (CAMELYON16).**

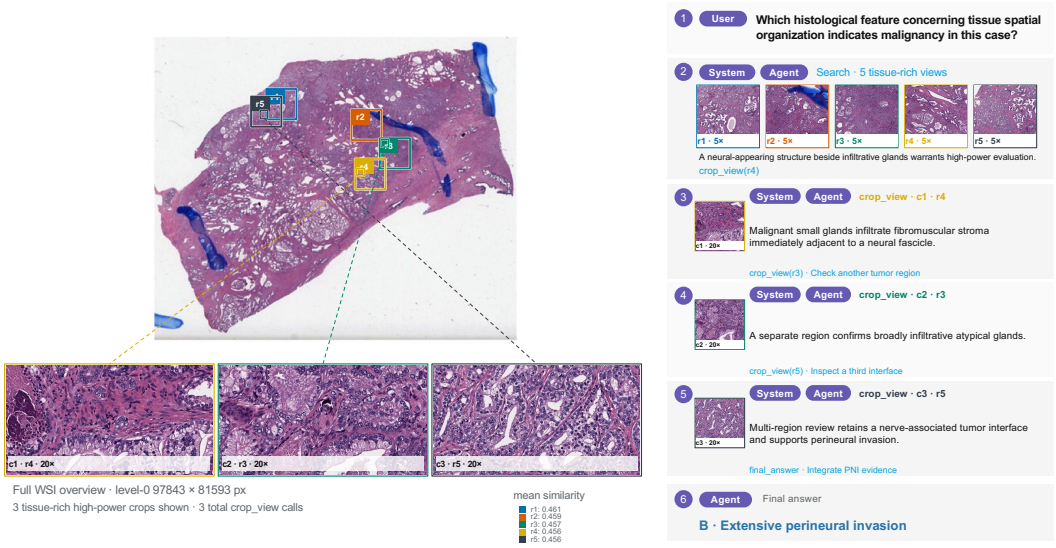

Figure 6: **Inspection of the interface between a nerve and tumor (SlideBench).**

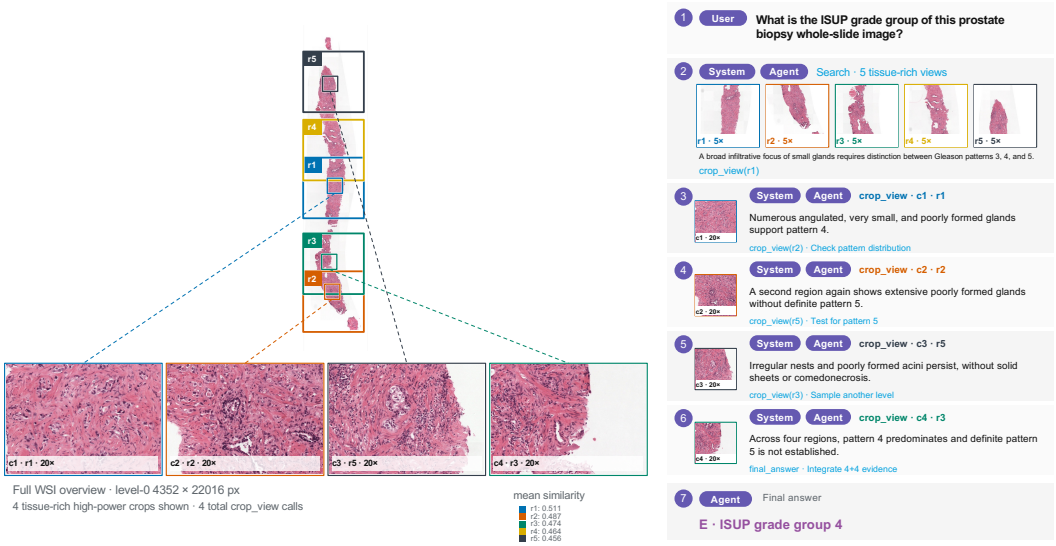

Figure 7: **Comparison of glandular architecture (PANDA).**

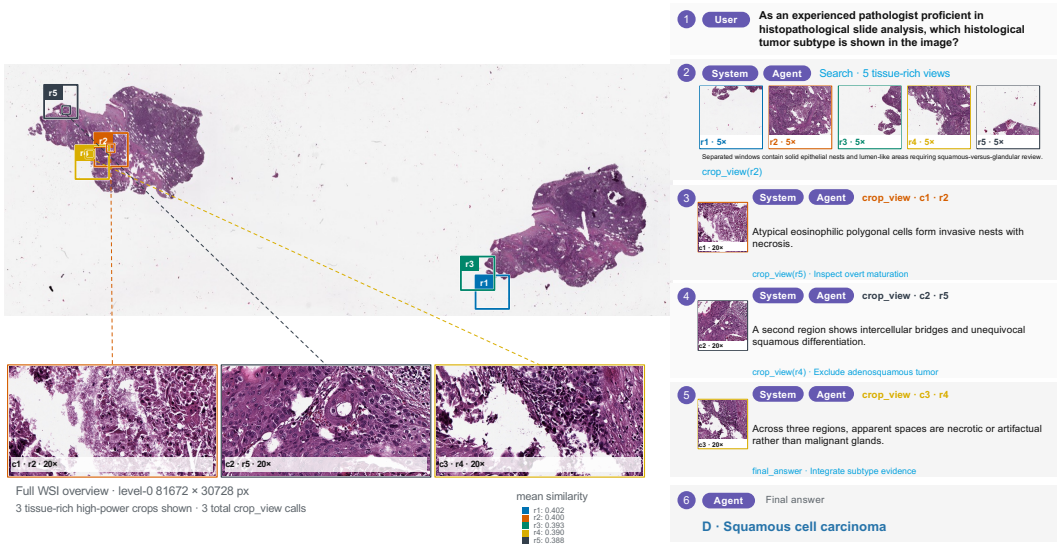

Figure 8: **Assessment of squamous differentiation (CPTAC-LSCC).**

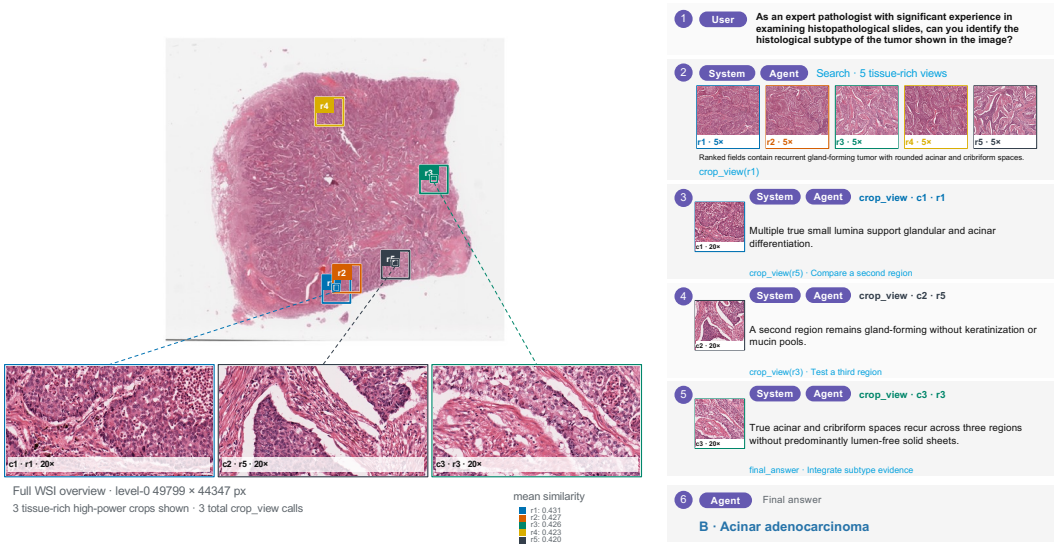

Figure 9: **Assessment of gland formation (CPTAC-LUAD).**

F PROMPT AND INTERFACE REFERENCE

This section explains the messages used for WSI interaction and training. The examples are formatted for readability while preserving the action names, fields, and output formats used by the system. The prompt templates and checker schemas are reproduced below.

F.1 POLICY PROMPT AND USER TASK

The policy is asked to begin with regional search, review the returned coarse views, inspect local tissue when finer morphology is useful, and answer when the available evidence is sufficient. A compact version of the instructions is:

```
You are a pathology visual reasoning agent operating on a
whole-slide image.

1. Start with navigate_with_keep. Provide 1-8 pathology keywords
   derived from the question and request up to five candidate views.
2. Review the returned coarse views.
3. If finer morphology is useful, call crop_view with a returned
   view_id and a bounding box in that view.
4. You may inspect additional regions or answer from the evidence
   already available.
5. Use only returned view identifiers and observed tissue evidence.
6. For a choice question, place one allowed option label in <answer>.
```

The first function corresponds to `Search`, and `crop_view` corresponds to `Crop_view`. In the prompt, "candidate views" denotes the $5\times$ coarse views of candidate regions. The diagram's selected view identifier and bounding box correspond to the `view_id` and `bbox_2d` fields below. The policy can answer directly after `Search` when the coarse views are sufficient; local inspection is optional.

For a choice question, the user message provides the case identifier, question, and answer options:

```
Case {case_id}
Question: {question}
Question kind: choice
Options:
A) {option_1}
B) {option_2}
...
Allowed option letters: A, B, ...
```

Option labels extend beyond Z when needed. Open questions used during trajectory construction provide the question without a fixed option set.

F.2 TOOL ARGUMENTS AND OBSERVATIONS

Table 9 gives the arguments visible to the model for the two actions.

Table 9: **Search and Crop_view function arguments.**

Tool	Arguments	Interpretation
`Search`	`keywords`: 1–8 strings; `max_views`: integer 1–8	The keywords describe tissue evidence to retrieve. We use five views by default.
`Crop_view`	`view_id`; `bbox_2d`=$[x_0, y_0, x_1, y_1]$	The view is one coarse view returned in the current episode. The box is expressed in that 1024×1024 view.

The runtime names are `navigate_with_keep` and `crop_view`. A crop request is rendered in the model's function call format, for example:

```
<tool_call>
<function=crop_view>
<parameter=view_id>{returned_view_id}</parameter>
<parameter=bbox_2d>[x0, y0, x1, y1]</parameter>
</function>
</tool_call>
```

A `Search` observation records the query and ranked coarse views, including their identifiers and coordinates on the slide. A `Crop_view` observation records the coarse view and location from which the tissue was read. Each returned image is inserted into the next model context alongside its observation record.

F.3 Assistant Messages and SFT Supervision

An assistant turn contains reasoning followed by a tool request or the terminal answer. The Qwen3.5 chat template renders reasoning inside a thinking block and tool calls in the structured representation above. SFT supervises the reasoning, tool requests, interpretation of returned observations, and terminal answer. Tool observations and returned images become input to later turns.

The same message order is used during online rollout. A Search action may be followed by zero or more Crops before the answer. Default evaluation uses the same interaction limits as training. In the call budget ablation, reaching the imposed limit restricts the next model turn to a terminal response.

F.4 Trajectory Quality Checking

The trajectory checker reads the question, reference answer, candidate answer, and the recorded `Search`/`Crop_view` evidence. It scores three aspects of the interaction:

Table 10: **Trajectory quality rubric.**

Field	Meaning
Correctness	Whether the final conclusion agrees with the reference answer
Completeness	Whether the response covers the information requested by the question
Grounding	Whether factual and visual claims are supported by the question and recorded observations
Unsupported claims	Material claims that are not supported by the available evidence
Explanation	A short comparison of the candidate answer, reference, and evidence

Correctness, completeness, and grounding use integer scores from 1 to 3. A concise answer can receive full completeness when the question does not require an explanation. The checker assesses whether the available observations support the answer and whether further local inspection is needed. For structured choice questions, the final option label is also checked deterministically against the gold answer.

Tool validity is checked separately from semantic quality. These checks cover action order, call/observation pairing, view identifiers, bounding boxes, and the presence of returned image artifacts.

F.5 Terminal Answers

A terminal response uses a short reasoning block followed by the answer:

```
<think>
brief evidence synthesis
</think>
<answer>answer only</answer>
```

For choice questions, the answer block contains one allowed option label. The evaluation parser checks the response format and then compares the predicted label with the gold label.

F.6 Fixed Question Templates

BRACS, CAMELYON16, and CAMELYON17 use fixed question and option templates. Dataset labels map directly to the corresponding answer choices.

Table 11: **Fixed questions for BRACS, CAMELYON16, and CAMELYON17.**

Dataset	Question	Answer space
BRACS	Which BRACS diagnostic category best describes this breast whole-slide image?	Normal; pathological benign lesion; UDH; FEA; ADH; DCIS; invasive carcinoma
CAMELYON16	Does this axillary lymph node whole-slide image contain metastatic breast carcinoma?	No metastatic carcinoma; metastatic carcinoma
CAMELYON17	What is the metastasis category in this axillary lymph node whole-slide image?	Negative; isolated tumor cells; micrometastasis; macrometastasis

F.7 Prompt Templates

The templates below preserve the roles, action names, fields, and answer formats used in the pipeline while removing repeated runtime warnings.

F.7.1 Policy system prompt

```
You are a pathology visual reasoning agent operating on a whole-slide image.

Use the provided tools to gather evidence from the slide.

1. Begin with navigate_with_keep. Generate 1-8 pathology keywords
   from the question and request up to five candidate views.
2. Review the returned coarse views.
3. When finer morphology is useful, call crop_view with a returned
   view_id and a bounding box in that view.
4. You may inspect more than one region, or answer directly if the
   current evidence is sufficient.
5. Base the final answer on the tissue evidence observed during the
   interaction.
6. For a choice question, return one allowed option label inside
   <answer>...</answer>.

Intermediate actions use the normal function-call format. A terminal
response uses:
<think>
brief evidence synthesis
</think>
<answer>answer only</answer>
```

F.7.2 Trajectory collection prompt

Trajectory collection before SFT uses the same visual actions and records each step as a compact JSON object:

```
You are a pathology visual reasoning agent operating on a whole-slide image.

Return one JSON object:
{
  "thought": "brief evidence summary",
  "action": {
    "tool": "navigate_with_keep | crop_view | final_answer",
    "param": {}
  },
  "plan": ["next intended step"]
}

Interaction:
- Start with navigate_with_keep using 1-8 pathology keywords.
- Use crop_view with a returned view_id and bbox_2d when local detail
  is useful.
- You may inspect several returned views.
- Use final_answer when the available evidence is sufficient.
- For a choice question, the answer is one allowed option label.
```

The JSON representation is used during SFT trajectory collection; the deployed policy uses the function call format described above.

F.7.3 USER TASK AND TERMINAL MESSAGE

```
Case {case_id}
Question: {question}
Question kind: choice
Options:
A) {option_1}
B) {option_2}
...
Allowed option letters: A, B, ...

No further tool calls are available. Answer using the evidence already
collected.

<think>
brief evidence synthesis
</think>
<answer>answer only</answer>
```

F.7.4 TRAJECTORY QUALITY CHECKER

The checker reads the question, reference, prediction, and the recorded interaction evidence. A compact version of the scoring prompt is:

```
Evaluate one pathology VQA answer.

Question:
{question}

Options:
{lettered_options_or_none}

Reference answer:
{reference_answer}

Candidate answer:
{prediction}

Recorded tool evidence:
{compact_json_evidence_log}

Score each item from 1 to 3:
- correctness: agreement with the reference answer
- completeness: whether the requested information is covered
- grounding: whether claims are supported by the question and
  recorded observations

Also list any material unsupported claims and give a short explanation.

A concise answer may be complete when the question only asks for a
label or short response. A crop is not required when the coarse
views already provide sufficient evidence.

Return JSON with:
correctness, completeness, grounding, unsupported_claims, explanation.
```

For closed questions, the predicted option label is also checked directly against the gold label. The semantic checker is mainly used to assess the quality of the recorded trajectory and to evaluate open answers.

F.7.5 ANSWER CORRECTNESS CHECKER

Open answers can also be evaluated with a smaller correctness prompt:

```
Judge whether the candidate answer correctly answers the reference.

Question:
{question}

Reference answer:
{reference_answer}
```

```
Candidate answer:
{prediction}

Return:
- correctness: 3 for correct, 2 for mostly correct or an acceptable
  paraphrase, 1 for incorrect
- explanation: a short justification
```

F.7.6 CHECKER OUTPUT SCHEMAS

The trajectory quality checker returns the following JSON object:

```
{
  "type": "object",
  "additionalProperties": false,
  "required": [
    "correctness",
    "completeness",
    "grounding",
    "unsupported_claims",
    "explanation"
  ],
  "properties": {
    "correctness": {"type": "integer", "minimum": 1, "maximum": 3},
    "completeness": {"type": "integer", "minimum": 1, "maximum": 3},
    "grounding": {"type": "integer", "minimum": 1, "maximum": 3},
    "unsupported_claims": {
      "type": "array",
      "maxItems": 32,
      "items": {"type": "string"}
    },
    "explanation": {
      "type": "string",
      "minLength": 1,
      "maxLength": 4000
    }
  }
}
```

The answer correctness checker uses a smaller schema:

```
{
  "type": "object",
  "additionalProperties": false,
  "required": ["correctness", "explanation"],
  "properties": {
    "correctness": {
      "type": "integer",
      "minimum": 1,
      "maximum": 3
    },
    "explanation": {
      "type": "string",
      "minLength": 1,
      "maxLength": 4000
    }
  }
}
```