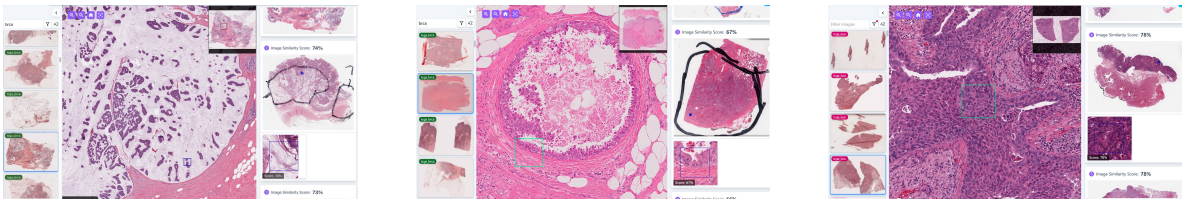
A Background

In the field of pathology, Whole Slide Imaging (WSI) has revolutionized the way tissue samples are analyzed, enabling digitization of high-resolution images for better diagnostic accuracy and research applications. However, the vast size and complexity of WSI data present significant challenges for manual inspection. Region retrieval in this context refers to the process of identifying and extracting specific regions within a WSI that are of diagnostic or research significance, such as areas exhibiting abnormal tissue patterns, tumors, or other pathological features. The need for such retrieval methods is particularly urgent given the increasing volume of WSI data and the growing demand for automated diagnostic tools that can aid pathologists in identifying critical regions swiftly. By automating this retrieval process, the workflow can be significantly accelerated, allowing pathologists to focus on higher-level analysis, thus improving the overall speed and accuracy of diagnosis.



(a) example 1 (b) example 2 (c) example 3

Figure 1: The examples of patch retrieval

Fig. 1 shows the limitations of current slide/patch retrieval in computational pathology:

- Fig. 1a demonstrates two critical limitations of patch-level retrieval systems in histopathological analysis. First, when querying with a breast cancer tissue patch, the system erroneously retrieves histologically similar patches from digestive system slides, despite their different organ origins. This reveals a fundamental flaw in current retrieval methodologies, which prioritize patch-level visual similarity while overlooking critical slide-level contextual information.
- The second limitation concerns diagnostic criteria requiring regional proportion analysis, as exemplified by mucinous carcinoma identification. Current patch-based approaches may retrieve isolated mucinous regions from non-malignant contexts rather than complete mucinous carcinoma slides, failing to meet clinical diagnostic requirements that depend on quantifying mucinous component percentages within entire tissue sections.
- Fig. 1b illustrates the system's inadequacy in handling large-scale pathological features, as exemplified by ductal carcinoma in situ (DCIS). Patch-based retrieval returns visually similar cellular patterns from non-DCIS cases, demonstrating the method's inability to recognize architectural features characteristic of DCIS that manifest at the whole-slide level rather than isolated patch regions.
- Fig. 1c reveals cross-category confusion between squamous cell carcinoma (lung origin) and adenocarcinoma (digestive system origin) at the patch level. While these distinct carcinoma types exhibit overlapping cytological features in isolated cellular regions, their diagnostic differentiation relies on slide-level architectural patterns that current patch-based systems cannot capture.

These limitations highlight the necessity for developing advanced region retrieval methodologies that incorporate: Multi-scale feature integration, Spatial relationship analysis, Context-aware similarity metrics, and Diagnostic pattern recognition. Such region-level retrieval systems would enable more clinically relevant histopathology analysis by preserving critical diagnostic context while maintaining cellular-level discriminative power, thereby better supporting precision medicine applications in digital pathology.

B Region Retrieval Task Definition

The goal of WSI region retrieval is to identify the region with the most similar semantic structure from a WSI dataset for a given query image, regardless of rotation or scale. A common approach to measuring semantic similarity between regions involves comparing their pixel-level segmentation masks. However, in practical settings, obtaining perfect segmentation masks is often infeasible, as complete semantic information in pathology images can only be extracted using foundation image encoders. To analyze WSI region retrieval more precisely, we first define the task using pixel-level segmentation masks and then reformulate it using an image encoder within the tessellation structure. Tab.1 summarizes the notations in this work.

Table 1: Summary of main notations.

Symbol	Description
$\mathcal{I}$	WSI set, a collection of WSIs
I	A single WSI
n	Number of WSIs
W, H	WSI dimensions
$D(I, l)$	Downsampling function with level $l \in [1, L]$
$Seg(\cdot)$	Segmentation model
M_l	Semantic mask with level l
d	Semantic space dimension
r_ϕ	Region of WSI based descriptor $\phi = (x, y, w, h, \theta)$
m_ϕ	Region of semantic mask based descriptor $\phi = (x, y, w, h, \theta)$
$f(a, b\|\cdot)$	Semantic density function
$Sim(\cdot, \cdot)$	Region Cosine similarity
$\alpha_{(\phi_1, \phi_2)}$	Aspect ratio difference
s	Semantic granularity ratio
M_l^s	Semantic mask with granularity ratio s
$E_{s^*}(\cdot)$	Image semantic encoder with minimum s^*
e_ϕ	Semantic embedding of region r_ϕ
$L^s(\cdot)$	Global function in $E_{s^*}(\cdot)$
$T_\phi^t = (V_\phi, R_\phi)$	Semantic tessellation with Vertex and relations
t	Sampling step
$p(e) = (x_e, y_e)$	affine identifier
$\Delta\theta, \Delta d$	The angular and length ratio shift of affine identifier
$R_p(\cdot\|\mathcal{I}, E_{s^*}(\cdot))$	Patch retrieval function
$\{\Delta\theta\}$	Angle shift set
$\{\Delta d\}$	Length ratio shift set
δ	Degree of coincidence
α	Correlation coefficient of semantic bound
ψ	Observed similarity of anchor
$\overline{sim}, \underline{sim}$	Similarity upper and lower bounds
$\mathcal{S}$	Square area in tessellation
ξ	Statistical error
$\text{dist}(i, j)$	distance of vertex in tessellation
$\text{sim}(i, j)$	similarity between vertexes of tessellation

B.1 Region Retrieval with Segmentation Mask

First, we define the WSI set, the pathology image, the semantic mask, and the corresponding region:

Definition 1. ***(WSI Set)*** *A WSIs set $\mathcal{I} = \{I\}^n$ is a collection of different WSI I with size $W \times H$, where:*

- *n denotes the total number of WSI in $\mathcal{I}$.*
- *I is a gigapixel image based on the WSI scanner magnification.*

Definition 2. ***(Pathology Image)*** *Given a WSI I, a pathology image $I_l \in \mathbb{R}^{W_l \times H_l \times 3}$ is an RGB image, which is determined by a down-sampling function D with I and sampling level $l \in [1, L]$:*

$$I_l = D(I, l), \text{ where } W_l = \lfloor W \times (l/L) \rfloor, H_l = \lfloor H \times (l/L) \rfloor,$$

where L is the upper bound of down-sampling, which is the maximum magnification of a micro scanner.

In particular, most WSI file formats (such as .svs and .tiff) are constructed using a pyramid structure with fixed indexes $[0, 1, 2, 3]$, which corresponds to down-sampling levels $l \in \{40, 20, 10, 5\}$ with $L = 40$, for faster visualization.

Definition 3. ***(Ideal Semantic Mask)*** *Given a pathology image I_l and an ideal segmentation model $Seg(\cdot)$, an ideal semantic mask M_l is a pixel-level mask corresponding to I_l, which is:*

$$M_l = Seg(I_l) \in \mathbb{R}^{W_l \times H_l \times d},$$

where d is the dimension of the total semantic space, and $M_l(x, y) \in \mathbb{R}^d$ is a sparse vector, which contains all the semantic information of point (x, y) at the pixel level based on I_l.

Definition 4. ***(Region)*** *Given a pathology image I_l or a semantic mask M_l, a WSI region $r_\phi \subseteq I_l$ or a semantic region $m_\phi \subseteq M_l$ is a rectangular area, which is described by an ordered quintuple descriptor $\phi = (x, y, w, h, \theta)$ with:*

$$r_\phi = I_l(\phi) \in \mathbb{R}^{w \times h \times 3}, \quad m_\phi = M_l(\phi) \in \mathbb{R}^{w \times h \times d}, \tag{1}$$

where (x, y) and (w, h) are the center position and size of regions r_ϕ in I_l or m_ϕ in M_l, respectively, with a rotation angle $\theta \in [-\pi, \pi)$. Here, $x, y, w, h > 0$, with $x - w/2, x + w/2 \in [0, W_l]$ and $y - h/2, y + h/2 \in [0, H_l]$.

Based on these definitions, the region retrieval is to find the most similar region (blue region in Fig. 1) from the existing $\mathcal{I}$ based on the query region r_{ϕ_q} (green region in Fig. 1) with only image information. To rigorously compare the information of regions with different sizes, we define the semantic density function and region similarity.

Definition 5. ***(Region Semantic Function)*** *Given a m_ϕ with $\phi = (x, y, w, h, \theta)$, a semantic density function $f(a, b|m_\phi, t)$ with a step t is a continuous function, which is:*

$$f(a, b|m_\phi, t) = m_\phi(i, j) \in \mathbb{R}^d,$$

where $\frac{i-t/2}{w} \le a < \frac{i+t/2}{w}$, $\frac{j-t/2}{h} \le b < \frac{j+t/2}{h}$, $i \in [0, w)$, $j \in [0, h)$.

Definition 6. ***(Region Similarity)*** *With two descriptor ϕ_1 and ϕ_2, the pixel-level similarity $Sim(\cdot, \cdot)$ between region r_{ϕ_1} and r_{ϕ_2} can be calculated with $f(a, b|m_{\phi_1}, 1)$ and $f(a, b|m_{\phi_2}, 1)$ by:*

$$Sim(r_{\phi_1}, r_{\phi_2}) = \alpha_{(\phi_1, \phi_2)} \iint_{\mathcal{D}} \frac{f(a, b|m_{\phi_1}, 1) \cdot f(a, b|m_{\phi_2}, 1)}{\|f(a, b|m_{\phi_1}, 1)\| \|f(a, b|m_{\phi_2}, 1)\|} \, da \, db,$$

where, $\alpha_{(\phi_1, \phi_2)} = exp\left(-|w_1 - w_2| - |h_1 - h_2|\right)$ is the aspect ratio difference between r_{ϕ_1} and r_{ϕ_2}, and $\mathcal{D} = [0, 1] \times [0, 1]$. Specifically, $Sim(r_{\phi_1}, r_{\phi_2}) = 1$ if and only if $m_{\phi_1} = m_{\phi_2}$, which called complete similarity and denoted as $r_{\phi_1} \sim r_{\phi_2}$.

Given an ideal segmentation model, Def.5 and Def.6 give how to measure the similarity between arbitrary regions. Based on the definition of region and region similarity, we define pixel-level region retrieval:

Definition 7. ***(Region Retrieval with Segmentation Mask)*** *For given query region r_{ϕ_q} with unknowable I_l, x, y and θ, the region retrieval task is to return a region $r_{\phi_{res}} \in \mathcal{I}$, which:*

$$\forall\, r_\phi \in \mathcal{I},\ s.t.\ Sim(r_{\phi_q}, r_{\phi_{res}}) \ge Sim(r_{\phi_q}, r_\phi), \tag{2}$$

where $\mathcal{I}$ is target WSI set.

B.2 Region Retrieval with Image Encoder

In many situations, obtaining a perfect segmentation semantic mask is challenging, since it requires an ideal segmentation model. To address this situation, current research primarily leverages foundational image encoders in pathology to approximate the semantic representation with patch-level regions in WSI. Specifically, we first redefine the region semantic mask and then define the image encoder:

Definition 8. ***(Semantic Mask)*** *Given a pathology image $I_l \in \mathbb{R}^{W_l \times H_l \times 3}$, the semantic mask $M_l \in \mathbb{R}^{W_l \times H_l \times S \times d_s}$ consists of information of I_l on semantic space $\mathbb{R}^{d_s}$ with the semantic granularity ratio $s \in [0, S]$, and $M_l(x, y, s) \subset M_l$ which represent the semantic information of r_ϕ with $\phi = (x, y, 2s, 2s, \theta)$ and arbitrary θ. The semantic region $m_\phi^s \subseteq M_l(s)$ with $\phi = (x, y, w, h, \theta)$.*

Definition 9. ***(Image Encoder)*** *Given a pathology image I_l and corresponding M_l, a model $E_{s^*}(r_\phi)$ is a image encoder, if $\forall\ \phi = (x, y, 2s, 2s, \theta),\ s \geq s^*$, $\exists$ global function $L^s(\cdot) : \mathbb{R}^{2s \times 2s} \to \mathbb{R}^{d_s}$ s.t.*

$$E_{s^*}(r_\phi) = L^s(m_\phi) = e_\phi \in \mathbb{R}^{d_s}, \tag{3}$$

where s^* is the minimum semantic granularity ratio supported by $E_{s^*}(r_\phi)$, and d_s is the size of semantic embedding e_ϕ.

Def.9 shows that the representation e_ϕ of the image encoder serves as the link between r_ϕ and m_ϕ, and its performance depends on s^*. However, the pixel-level segmentation semantic mask $M_l \in \mathbb{R}^{W_l \times H_l \times 1 \times d_1}$ is typically not fully captured by the current encoder. This is because the encoder, with the scale s^*, often fails to capture semantic information in regions smaller than $s < s^*$.

In practice, we propose a region tessellation method that generalizes the semantic region representation m_ϕ at a specific granularity level s, defined as follows.

Definition 10. ***(Region Tessellation)*** *Given a region r_ϕ with $\phi = (x, y, w, h, \theta)$ and a pathology image encoder $E_{s^*}(\cdot)$, the region tessellation $T_\phi^t = \{V_\phi, R_\phi\}$ is an overlapping square tiling composed of sampled semantic region anchors m_{ϕ_g} with a step size t and their adjacency relations:*

$$\begin{aligned} V_\phi &= \{(m_{\phi_g}, x, y) \mid \phi_g = (x_g, y_g, 2s, 2s, \theta), m_\phi(x_g, y_g, s)\}, \\ R_\phi &= \{(m_{\phi_u}, m_{\phi_v}) \mid m_{\phi_u}, m_{\phi_v} \in V_\phi,\ Adj(m_{\phi_u}, m_{\phi_v})\}, \end{aligned} \tag{4}$$

where $T_\phi^t(x_g, y_g) = m_{\phi_g}$, $(x_g, y_g) = (x + x' \cdot \cos(\theta) - y' \cdot \sin(\theta),\ y + x' \cdot \sin(\theta) + y' \cdot \cos(\theta))$ is the center point of anchor and $x' \in [-w/2, w/2]$, $y' \in [-h/2, h/2]$, $Adj(m_{\phi_u}, m_{\phi_v}) = \{|x_u - x_v| = t\} \oplus \{|y_u - y_v| = t\}$ is adjacency condition function, and the sampling with step size t is performed along the vertical and horizontal axes of the region r_ϕ.

The region tessellation consists of anchor points sampled from m_ϕ^s using a predefined step size t. This structure captures the relative spatial relationships among anchors within the semantic mask, enabling region representations that are invariant to rotation and scaling. Specifically, given the encoder output $E_s(r_{\phi'})$, we define the semantic mask m_ϕ^s as T_ϕ^1 at granularity level s. Note that T_ϕ^1 corresponds to a pixel-level segmentation mask when $s = 1$. According to Def. 6 and Def. 10, the similarity between two region tessellations is defined as:

Definition 11. ***(Region Tessellation Similarity)*** *Given two region r_{ϕ_1} and r_{ϕ_2} and an image encoder $E_{s^*}(\cdot)$, the region tessellation similarity $Sim_T(\cdot, \cdot)$ between r_{ϕ_1} and r_{ϕ_2} can be calculated with $T_{\phi_1}^t$ and $T_{\phi_2}^t$ by:*

$$Sim_T(r_{\phi_1}, r_{\phi_2}) = \alpha_{(\phi_1, \phi_2)} \iint_{\mathcal{D}} \frac{f(a, b|T_{\phi_1}^t, t) \cdot f(a, b|T_{\phi_2}^t, t)}{\|f(a, b|T_{\phi_1}^t, t)\| \|f(a, b|T_{\phi_2}^t, t)\|}\, da\, db,$$

where, $\alpha_{(\phi_1, \phi_2)} = exp\,(-|w_1 - w_2| - |h_1 - h_2|)$ is the aspect ratio difference between r_{ϕ_1} and r_{ϕ_2}, and $\mathcal{D} = [0, 1] \times [0, 1]$.

In particular, $Sim_T(r_{\phi_1}, r_{\phi_2})$ is equivalent to $Sim(r_{\phi_1}, r_{\phi_2})$ with semantic ratio s^* if $t = 1$. Based on the Def.10 and Def.11, we here give the new definition of the region retrieval task:

Definition 12. ***(Region Retrieval with Tessellation)*** *For given query region r_{ϕ_q} with unknowable I_l, x, y, θ and encoder $E_{s^*}(\cdot)$, the region retrieval task is to return a region $r_{\phi_{res}} \in \mathcal{I}$, which $\forall\ r_\phi \in \mathcal{I}$, $s \geq s^*$ and $t \leq s$, s.t.:*

$$Sim_T(r_{\phi_q}, r_{\phi_{res}}) \geq Sim_T(r_{\phi_q}, r_\phi), \tag{5}$$

where $\mathcal{I}$ is target WSI set.

B.3 Coordinate Convention and Valid Region Constraints

For all equations in this appendix, we use the same image coordinate frame as the main paper: the origin is at the top-left corner of I_l, the x-axis points to the right, and the y-axis points downward. Rotation angle θ is measured in radians with principal range $[-\pi, \pi)$.

Given $\phi = (x, y, w, h, \theta)$, a region is valid only if all transformed corner coordinates remain within image bounds. In implementation, we apply two practical rules:

- If a candidate region slightly exceeds the image boundary, we clip it to the largest valid rectangle and keep its center and angle unchanged.
- If clipping removes too much support (less than 70% valid area), the candidate is discarded from re-ranking.

To avoid ambiguity in angle arithmetic, all angle differences are normalized by

$$\text{wrap}(\Delta\theta) = ((\Delta\theta + \pi) \bmod 2\pi) - \pi.$$

B.4 Sampling Geometry in Tessellation

For tessellation step t, anchors are sampled on the local region axes before global rotation:

$$x' = -\frac{w}{2} + it, \quad y' = -\frac{h}{2} + jt,$$

where i, j are integers such that sampled anchors remain in the valid support. The corresponding global center is obtained by the same rigid transform used in Def. 10.

In practice, we denote the effective anchor count by

$$N_a \approx \left\lfloor \frac{w}{t} \right\rfloor \left\lfloor \frac{h}{t} \right\rfloor,$$

and use 4-neighborhood adjacency at stride t for relation construction. This keeps graph connectivity stable under moderate rotation and scale changes while limiting query-time complexity.

C Property

Assumptions. We assume: (i) anchor pairs used for affine estimation are non-degenerate ($\|p\| > 0$), (ii) both regions are represented in the same 2D coordinate frame, and (iii) angle differences are normalized by wrap$(\cdot)$ into $[-\pi, \pi)$.

Property 1. *Given a region r_ϕ with $\phi = (x, y, w, h, \theta)$ and one edge vector $p = (x_v - x_u, y_v - y_u)$ between two anchors, let $r_{\phi'}$ be the transformed region with rotation increment $\Delta\theta$ and isotropic scale $d > 0$. Then:*

$$p' = d\, R(\Delta\theta)\, p,$$

where $R(\Delta\theta)$ is the 2D rotation matrix. Consequently, the orientation shift equals $\Delta\theta$ (modulo wrap-around) and the length ratio equals d.

Proof. Let

$$R(\Delta\theta) = \begin{bmatrix} \cos\Delta\theta & -\sin\Delta\theta \\ \sin\Delta\theta & \cos\Delta\theta \end{bmatrix}.$$

For anchor coordinates $u = (x_u, y_u)$ and $v = (x_v, y_v)$ in the local frame, the transformed coordinates satisfy

$$u' = dR(\Delta\theta)u + b, \qquad v' = dR(\Delta\theta)v + b,$$

where b is translation. Subtracting the two equations removes translation:

$$p' = v' - u' = dR(\Delta\theta)(v - u) = dR(\Delta\theta)p.$$

Taking norms gives scale consistency:

$$\|p'\| = d\|p\| \Rightarrow \Delta d = \frac{\|p'\|}{\|p\|} = d.$$

For orientation, if $\angle(p)$ denotes the directed angle of p, then

$$\angle(p') = \angle(dR(\Delta\theta)p) = \angle(p) + \Delta\theta.$$

Hence

$$\text{wrap}(\angle(p') - \angle(p)) = \Delta\theta.$$

In practice, we compute the unsigned part with

$$\arccos\left(\frac{p \cdot p'}{\|p\|\|p'\|}\right),$$

and recover sign by the 2D cross product term $p_x p'_y - p_y p'_x$ (equivalently by atan2). □

Corner cases. If $\|p\| = 0$ or $\|p'\| = 0$, angle is undefined and the pair is discarded. When the angle is near $\pm\pi$, we always apply wrap normalization to avoid discontinuity.

D Algorithm

To operationalize the aforementioned formulation, we design the **URICA** region retrieval algorithm (Alg. 1), which follows a two-stage process: (1) *affine identifier estimation* and (2) *region reconstruction and ranking*.

D.1 Parameter Definition

The algorithm uses the following parameters:

- t: tessellation stride for anchor sampling.
- k_a: number of top anchor retrievals per query anchor used for voting and affine estimation.
- k: final number of returned region candidates.
- τ_θ, τ_d: inlier thresholds for angle and scale consistency filtering.

Unless otherwise stated, all anchors and candidate regions are encoded by the same backbone $E_{s^*}(\cdot)$ and compared by cosine similarity.

Algorithm 1 URICA Region Retrieval Algorithm

Input: Query region r_{ϕ_q}, target WSI set $\mathcal{I}$, encoder $E_{s^*}(\cdot)$, step size t, patch retrieval size k_a, final return size k, inlier thresholds τ_θ, τ_d
Output: Top-k region candidates in $\mathcal{I}$
Build query tessellation $T^t_{\phi_q} = \{V_{\phi_q}, R_{\phi_q}\}$
for each anchor $v^*_j \in V_{\phi_q}$ **do**
    Compute anchor embedding $e^*_j = E_{s^*}(v^*_j)$
    Retrieve top-k_a anchors $\mathcal{R}_j = R_{\mathcal{I}}(e^*_j \mid \mathcal{I}, E_{s^*}, k_a)$
    Vote target slide/location from $\mathcal{R}_j$
end for
Keep the most voted slide-location pair (I^*, l^*)
for each matched anchor pair (v^*_u, v^*_v) with candidates in (I^*, l^*) **do**
    Build pointers $p = v^*_u \to v^*_v$ and $p_{res} = \hat{v}_u \to \hat{v}_v$
    Estimate $\Delta\theta = \angle(p, p_{res})$ and $\Delta d = \|p_{res}\|/\|p\|$
    if $|\Delta\theta - \text{med}(\Delta\theta)| \le \tau_\theta$ and $|\Delta d - \text{med}(\Delta d)| \le \tau_d$ **then**
        Add $(\Delta\theta, \Delta d)$ into inlier set
    end if
end for
Compute robust affine estimates $\theta^* = \text{mean}(\{\Delta\theta\}_{inlier})$, $d^* = \text{mean}(\{\Delta d\}_{inlier})$
for each anchor hypothesis in (I^*, l^*) **do**
    Reconstruct candidate descriptor $\phi_{res} = (x_{res}, y_{res}, w_q \cdot d^*, h_q \cdot d^*, \theta_q + \theta^*)$
    Crop candidate region $r_{\phi_{res}}$ and append to candidate list
end for
for each candidate region $r_{\phi_{res}}$ **do**
    Compute query-candidate similarity with $E_{s^*}(\cdot)$
end for
Sort all candidates by similarity and return top-k

In the first stage, URICA iterates over each anchor v_{ϕ^*} in the query tessellation $T^t_{\phi_q}$. Each patch is encoded by the pretrained encoder $E_{s^*}(\cdot)$ to obtain an embedding $e_\phi = E_{s^*}(v_\phi)$, which retrieves the top-k most similar patches $\{r_{\phi^*_{res}}\}^k = R_{\mathcal{I}}(e^*_\phi | \mathcal{I}, E_{s^*})$. A voting process aggregates retrieval frequencies across WSIs and patch locations to determine the most likely target slide I^* and anchor position l^*.

In the second stage, for each anchor pair $\{v_{\phi^*}, v_{\phi'}\}$, URICA retrieves their corresponding results $\{r_{\phi^*_{res}}\}^k$ and $\{r_{\phi'_{res}}\}^k$ and computes local affine shifts. Given geometric pointers $p : v_{\phi^*} \to v_{\phi'}$ and $p_{res} : r_{\phi^*_{res}} \to r_{\phi'_{res}}$, rotation and scale differences are estimated as:

$$\Delta\theta = \arccos \frac{p \cdot p_{res}}{\|p\|\|p_{res}\|}, \quad \Delta d = \frac{\|p_{res}\|}{\|p\|}. \tag{6}$$

All $\{\Delta\theta\}$ and $\{\Delta d\}$ are aggregated, and their most consistent subsets $\{\Delta\theta\}^*$ and $\{\Delta d\}^*$ are selected by minimizing intra-set variance. The averaged parameters $\theta^* = \frac{1}{|\{\theta\}^*|}\sum_{\theta\in\{\theta\}^*}\theta$ and $d^* = \frac{1}{|\{d\}^*|}\sum_{d\in\{d\}^*} d$ define the dominant affine transformation.

The candidate region descriptor is then constructed as $\phi_{res} = (x_{res}, y_{res}, w_{res}, h_{res}, \theta_{res})$. Each candidate $r_{\phi_{res}}$ is evaluated via embedding similarity with $E_{s^*}(r_{\phi_q})$, and the top-k most similar regions are returned as the final results.

D.2 Complexity Analysis

Let $N_a = |V_{\phi_q}|$ be the number of query anchors, and let C_E denote one forward pass cost of $E_{s^*}(\cdot)$. Let $C_R(k_a)$ denote one ANN retrieval cost for top-k_a in the HNSW index.

Stage 1 (anchor encoding and voting). Encoding and retrieval over all anchors costs

$$\mathcal{O}(N_a \cdot (C_E + C_R(k_a))).$$

Stage 2 (affine estimation and reconstruction). Using N_p valid anchor pairs for affine estimation, the consistency filtering and robust aggregation cost

$$\mathcal{O}(N_p).$$

Candidate reconstruction and re-ranking over N_c region hypotheses cost

$$\mathcal{O}(N_c \cdot C_E + N_c \log N_c).$$

Total query complexity. Ignoring constant factors, total time is

$$\mathcal{O}(N_a(C_E + C_R(k_a)) + N_p + N_c C_E + N_c \log N_c).$$

The online memory footprint is dominated by query anchor embeddings and candidate buffers, i.e.,

$$\mathcal{O}(N_a d_s + N_c d_s),$$

while the database-side index memory is handled offline by Milvus/HNSW.

E Task Relationship

We analyze the upper and lower bounds of the similarity for an unobserved point (x, y) based on the observed anchor in a square region $\mathcal{S} = [0, t] \times [0, t]$ within a tessellation. Suppose the similarity of the observed anchors follows a uniform distribution: $\psi_i \sim U(0, 1) = 1$, where $\psi = [0, 1]$. We assume that, based on an observed anchor at (x^*, y^*), the upper and lower bounds of the similarity for an unobserved point (x, y) are related to the standard alignment between the two points, as follows:

$$\begin{aligned} \overline{sim}_i(x, y) =& \psi_i + (1 - \psi_i) \cdot (1 - \delta_i) = 1 - (1 - \psi_i)\delta_i && \in [\psi_i, 1], \\ \underline{sim}_i(x, y) =& \psi_i - \psi_i \cdot (1 - \delta_i) = \psi_i \delta_i && \in [0, \psi_i], \end{aligned}$$

where $\delta_i = \left(\left(1 - \frac{|x - x_i^*|}{2s}\right)\left(1 - \frac{|y - y_i^*|}{2s}\right)\right)^\alpha \in [0, 1]$, s is the semantic granularity radius determined by the image encoder, and $\alpha > 0$ is a coefficient used to adjust the effect of displacement on the similarity change. In particular, we have the following expressions for δ_i:

$$\begin{aligned} \delta_1 = \left(1 - \frac{x}{2s}\right)^\alpha \left(1 - \frac{y}{2s}\right)^\alpha, \quad \delta_2 = \left(1 - \frac{t - x}{2s}\right)^\alpha \left(1 - \frac{y}{2s}\right)^\alpha, \\ \delta_3 = \left(1 - \frac{x}{2s}\right)^\alpha \left(1 - \frac{t - y}{2s}\right)^\alpha, \quad \delta_4 = \left(1 - \frac{t - x}{2s}\right)^\alpha \left(1 - \frac{t - y}{2s}\right)^\alpha. \end{aligned}$$

It is noteworthy that the following property holds: $\delta_1 \delta_4 = \delta_2 \delta_3$.

Assumption clarification. In this section, $\psi_i \sim U(0, 1)$ means the random variable has probability density function $f_{\psi_i}(u) = 1$ for $u \in [0, 1]$ (not point probability). We also assume i.i.d. anchors and $0 < t < s$.

E.1 Boundary analysis

We define the upper bound of the similarity for any $(x, y) \in \mathcal{S}$ as the minimum of the similarity bounds for the four observed anchor points on the square, that is:

$$\overline{sim}(x, y) = \min_{i=1,2,3,4} \{\overline{sim}_i(x, y)\} \tag{7}$$

Based on the order statistics of ψ_i, we can compute:

$$P(\overline{sim}(x, y) = t) = \sum_{i=1}^{4} P(\overline{sim}_i(x, y) = t) \prod_{j \neq i} P(\overline{sim}_j(x, y) > t) \tag{8}$$

For $\overline{sim}_i(x, y) = t$, we have the relation: $1 - (1 - \psi_i) \cdot \delta_i = t \quad \Rightarrow \quad \psi_i = 1 - \frac{1-t}{\delta_i}$. Thus, we can write:

$$f_{\overline{sim}_i}(t) = f_{\psi_i}\left(1 - \frac{1 - t}{\delta_i}\right) \left|\frac{d\psi_i}{dt}\right| = \frac{1}{\delta_i}, \quad P(\overline{sim}_i(x, y) > t) = 1 - P\left(\psi_i \leq 1 - \frac{1 - t}{\delta_i}\right) = \frac{1 - t}{\delta_i}$$

Therefore, the expected value of the upper bound of similarity at a given point (x, y) is:

$$E(\overline{sim}(x, y)) = \sum_{i=1}^{4} \int t \cdot \prod_{j \neq i} \left(\frac{1 - t}{\delta_j}\right) dt \tag{9}$$

We consider the condition $1 > \psi_i > k$, which gives $1 > t > 1 - (1 - k)\delta_i$. For $i = 1$, we have:

$$\int_{1-(1-k)\delta_1}^{1} \frac{t(1 - t)^3}{\delta_2 \delta_3 \delta_4} dt = \frac{\frac{1}{4}(1 - k)^4 \delta_1^4 - \frac{1}{5}(1 - k)^5 \delta_1^5}{\delta_2 \delta_3 \delta_4} \tag{10}$$

Thus, the expected value of the upper bound of similarity is:

$$E(\overline{sim}(x, y)) = \frac{\frac{1}{4}(1 - k)^4 \sum_i \delta_i^5 - \frac{1}{5}(1 - k)^5 \sum_i \delta_i^6}{\delta_1 \delta_2 \delta_3 \delta_4} \tag{11}$$

Similarly, we analyze the lower bound of similarity. For any $(x, y) \in \mathcal{S}$, the lower bound of the similarity for (x, y) is the maximum of the similarity bounds estimated by the four observed anchor points on the square, that is:

$$\underline{sim}(x, y) = \max_{i=1,2,3,4} \{\underline{sim}_i(x, y)\} \tag{12}$$

Based on the order statistics, we can compute:

$$P(\underline{sim}(x, y) = t) = \sum_{i=1}^{4} P(\underline{sim}_i(x, y) = t) \prod_{j \neq i} P(\underline{sim}_j(x, y) < t) \tag{13}$$

For $\underline{sim}_i(x, y) = t$, we have the following relation: $\psi_i \cdot \delta_i = t \quad \Rightarrow \quad \psi_i = \frac{t}{\delta_i}$. Thus, we obtain:

$$f_{\underline{sim}_i}(t) = f_{\psi_i}\left(\frac{t}{\delta_i}\right)\left|\frac{d\psi_i}{dt}\right| = \frac{1}{\delta_i}, \quad P(\underline{sim}_i(x, y) < t) = P\left(\psi_i < \frac{t}{\delta_i}\right) = \frac{t}{\delta_i}$$

Therefore, the expected value of the lower bound of similarity at a given point (x, y) is:

$$E(\underline{sim}(x, y)) = \sum_{i=1}^{4} \int t \cdot \prod_{j \neq i} \left(\frac{t}{\delta_j}\right) dt \tag{14}$$

We consider the condition $1 > \psi_i > k$, which gives $\delta_i > t > k\delta_i$. For $i = 1$, we have:

$$\int_{k\delta_1}^{\delta_1} \frac{t^4}{\delta_2\delta_3\delta_4}\, dt = \frac{\frac{1}{5}(1 - k^5)\delta_1^5}{\delta_2\delta_3\delta_4} \tag{15}$$

Thus, the expected value of the lower bound of similarity is:

$$E(\underline{sim}(x, y)) = \frac{\frac{1}{5}(1 - k^5)\sum_i \delta_i^6}{\delta_1\delta_2\delta_3\delta_4} \tag{16}$$

E.2 Analysis

We can calculate the average interval length of similarity for the given anchor and for any $(x, y) \in \mathcal{S}$ as follows:

$$L(sim(x, y)) = E(\overline{sim}(x, y)) - E(\underline{sim}(x, y)) = \frac{\frac{1}{4}(1 - k)^4 \sum_i \delta_i^5 - \frac{1}{5}(1 - k)^5 \sum_i \delta_i^6 - \frac{1}{5}(1 - k^5)\sum_i \delta_i^6}{\delta_1\delta_2\delta_3\delta_4}$$

Consider the coefficient function: $f(k) = \frac{1}{4}(1 - k)^4 - \frac{1}{5}(1 - k)^5 - \frac{1}{5}(1 - k^5)$, the derivative of $f(k)$ is:

$$f'(k) = -(1 - k)^3 + (1 - k)^4 + k^4 = -k(1 - k)^3 + k^4 = k(k^3 - (1 - k)^3)$$

For high similarity ($k > 0.5$), the derivative is positive, and since $f(1) = 0$, we have $f(k) < 0$, i.e.,

$$\frac{1}{4}(1 - k)^4 < \frac{1}{5}(1 - k)^5 + \frac{1}{5}(1 - k^5)$$

Thus, we can simplify:

$$L(sim(x, y)) < \frac{\frac{1}{4}(1 - k)^4 \left(\delta_1^5 + \delta_2^5 + \delta_3^5 + \delta_4^5 - \delta_1^6 - \delta_2^6 - \delta_3^6 - \delta_4^6\right)}{\delta_1\delta_2\delta_3\delta_4} \tag{17}$$

Using the property $\delta_1\delta_4 = \delta_2\delta_3$, we have:

$$\frac{\frac{1}{4}(1 - k)^4 \left(\sum_{i=1}^4 \delta_i^5 - \sum_{i=1}^4 \delta_i^6\right)}{\delta_1\delta_2\delta_3\delta_4} = \frac{1}{4}(1 - k)^4 \left(\frac{\delta_1^3}{\delta_4^2} + \frac{\delta_4^3}{\delta_1^2} + \frac{\delta_2^3}{\delta_3^2} + \frac{\delta_3^3}{\delta_2^2} - \frac{\delta_1^4}{\delta_4^2} - \frac{\delta_4^4}{\delta_1^2} - \frac{\delta_2^4}{\delta_3^2} - \frac{\delta_3^4}{\delta_2^2}\right)$$

This represents the average interval length for a given point. Considering the term $\frac{\delta_1^3}{\delta_4^2}$, we have:

$$\iint_{\mathcal{S}} \frac{\delta_1^3}{\delta_4^2}\, ds = \int_0^t \int_0^t \frac{\left(1 - \frac{x}{2s}\right)^{3\alpha}\left(1 - \frac{y}{2s}\right)^{3\alpha}}{\left(1 - \frac{t-x}{2s}\right)^{2\alpha}\left(1 - \frac{t-y}{2s}\right)^{2\alpha}}\, dx\, dy = \left(\frac{1}{(2s)^\alpha}\int_0^t \frac{(2s - x)^{3\alpha}}{(2s - t + x)^{2\alpha}}\, dx\right)^2$$

Similarly, we can derive:

$$\iint_{\mathcal{S}} \frac{\delta_4^3}{\delta_1^2}\,ds = \left(\frac{1}{(2s)^{\alpha}}\int_0^t \frac{(2s-t+x)^{3\alpha}}{(2s-x)^{2\alpha}}\,dx\right)^2$$

$$\iint_{\mathcal{S}} \frac{\delta_2^3}{\delta_3^2}\,ds = \iint_{\mathcal{S}} \frac{\delta_3^3}{\delta_2^2}\,ds = \frac{1}{(2s)^{2\alpha}}\int_0^t \frac{(2s-t+x)^{3\alpha}}{(2s-x)^{2\alpha}}\,dx \int_0^t \frac{(2s-x)^{3\alpha}}{(2s-t+x)^{2\alpha}}\,dx$$

Using $d(t-x) = -dx$, we obtain:

$$\int_0^t \frac{(2s-x)^{3\alpha}}{(2s-t+x)^{2\alpha}}\,dx \overset{\text{let}}{=} \int_0^t A(x)\,dx = \int_0^t \frac{(2s-t+x)^{3\alpha}}{(2s-x)^{2\alpha}}\,dx$$

$$\int_0^t \frac{(2s-x)^{4\alpha}}{(2s-t+x)^{2\alpha}}\,dx \overset{\text{let}}{=} \int_0^t B(x)\,dx = \int_0^t \frac{(2s-t+x)^{4\alpha}}{(2s-x)^{2\alpha}}\,dx$$

Thus, we have:

$$\begin{aligned}\iint_{\mathcal{S}} L(sim(x,y))\,ds <& \frac{(1-k)^4}{(2s)^{2\alpha}}\left(\int_0^t A(x)\,dx\right)^2 - \frac{(1-k)^4}{(2s)^{4\alpha}}\left(\int_0^t B(x)\,dx\right)^2 \\ =& \frac{(1-k)^4}{(2s)^{2\alpha}}\left(\left(\int_0^t A(x)\,dx\right)^2 - \left(\int_0^t \frac{B(x)}{(2s)^{\alpha}}\,dx\right)^2\right)\end{aligned}$$

Considering $\alpha > 0$, $0 < t < s$, we have:

$$\begin{aligned}\frac{B(x)}{(2s)^{\alpha}} =& \frac{(2s-x)^{4\alpha}}{(2s)^{\alpha}(2s-t+x)^{2\alpha}} < \frac{(2s-x)^{3\alpha}}{(2s-t+x)^{2\alpha}} = A(x) \\ A'(x) =& -\alpha(2s-x)^{3\alpha-1}(2s-t+x)^{-2\alpha-1}(10s-3t+x) < 0 \\ B'(x) =& -\alpha(2s-x)^{4\alpha-1}(2s-t+x)^{-2\alpha-1}(12s-4t+2x) < 0\end{aligned}$$

Thus, $A(x)$ and $B(x)$ are monotonically decreasing functions over their domains, and:

$$\frac{t(2s-t)^{3\alpha}}{(2s)^{2\alpha}} < \int_0^t A(x)\,dx < \frac{t(2s)^{3\alpha}}{(2s-t)^{2\alpha}}, \quad \frac{t(2s-t)^{4\alpha}}{(2s)^{3\alpha}} < \int_0^t \frac{B(x)}{(2s)^{\alpha}}\,dx < \frac{t(2s)^{3\alpha}}{(2s-t)^{2\alpha}} \tag{18}$$

Thus, we have:

$$\begin{aligned}\iint_{\mathcal{S}} L(sim(x,y))\,ds <& \frac{(1-k)^4}{(2s)^{2\alpha}}\left(\left(\int_0^t A(x)\,dx\right) + \left(\int_0^t \frac{B(x)}{(2s)^{\alpha}}\,dx\right)\right)\left(\left(\int_0^t A(x)\,dx\right) - \left(\int_0^t \frac{B(x)}{(2s)^{\alpha}}\,dx\right)\right) \\ <& \frac{(1-k)^4}{(2s)^{2\alpha}}\left[\left(\frac{t(2s)^{3\alpha}}{(2s-t)^{2\alpha}} + \frac{t(2s)^{3\alpha}}{(2s-t)^{2\alpha}}\right) \times \left(\frac{t(2s)^{3\alpha}}{(2s-t)^{2\alpha}} - \frac{t(2s-t)^{4\alpha}}{(2s)^{3\alpha}}\right)\right] \\ =& 2t^2(1-k)^4\left[\frac{(2s)^{4\alpha}}{(2s-t)^{4\alpha}} - \frac{(2s-t)^{2\alpha}}{(2s)^{2\alpha}}\right]\end{aligned}$$

E.3 Practical Rule for Choosing

Given tolerance $\xi \in (0,1)$, we select t by requiring the expected interval length to remain small with high probability:

$$\Pr\left(\iint_{\mathcal{S}} L(sim(x,y))\,ds \le \eta\right) \ge 1-\xi.$$

Combining the upper bound above with query-time cost $\mathcal{O}(1/t^2)$ for dense sampling yields a practical trade-off: very small t increases cost with limited bound improvement, while large t weakens the bound. In our setting, this balance is achieved around $t \approx 60$ (with $\alpha = 3$ in Sec. F.2), which is therefore used as the default stride.

F Experiments

F.1 Experimental Setup Details

Dataset construction and cohort split. All experiments are conducted on a TCGA-based cohort containing 24,811 WSIs from 29 cancer subtypes and 10 anatomical sites. We follow the same site taxonomy used in the main paper: Pulmonary, Urinary, Melanocytic, Brain, Gastrointestinal, Liver/Pancreas-Biliary, Gynecologic, Prostate/Testis, Hematopoiesis, and Endocrine. WSIs with severe artifacts (blank tissue area ratio larger than 70%, scanner failure, or missing diagnostic metadata) are excluded before indexing.

Patch generation and magnification. For retrieval, each WSI is first accessed at thumbnail level for coarse localization, then region candidates are sampled at diagnostic magnification. Query and database anchors use 224×224 patches; region-level evaluation uses reconstructed windows with minimum size 448×448. Random rotations are sampled from $[-180°, 180°]$. For each site, 18,269 randomized region queries are generated near image trisections to avoid center-only bias.

Encoder and embedding setup. Unless otherwise noted, URICA uses the UNI encoder with fixed embedding dimension and cosine normalization. Anchor embeddings are precomputed offline for indexed WSIs. During query time, only query-side embeddings and candidate re-ranking embeddings are computed online.

Vector index and search backend. We use Milvus with an HNSW index for patch-level retrieval. The deployed index configuration is: metric = cosine, $M = 32$, $efConstruction = 200$, and query-time $efSearch = 128$. These settings are fixed for all methods that rely on ANN retrieval to ensure fair latency comparison.

Evaluation protocol. All metrics are reported as macro-averages over queries within each site and over the full cohort. Slide retrieval uses mMV@k. Region retrieval uses mSim@k, mIoU@k, and mean time per query (mTPQ). Confidence intervals are estimated by bootstrap resampling (1,000 rounds). Random seeds are fixed for query generation and sampling.

Slide Retrieval: Queries consist of anatomical site, input image, and cancer subtype. The goal is to retrieve slides matching the query subtype within the site. Performance is measured by modified majority voting (mMV@k):

$$\text{mMV@k} = \frac{1}{N}\sum_{i=1}^{N} \mathbb{I}[L_i = MV(ret_i[1:k])],$$

where N is query count, L_i the true label, and $MV(ret_i[1:k])$ the majority vote among top-k retrievals. URICA is compared to Yottixel, SISH, RetCCL, and HSHR using their best reported settings, with queries drawn from the full WSI thumbnails and metadata.

F.1.1 Region Retrieval

This task tests URICA's ability to find histologically similar regions across WSIs. Performance metrics include Sim@k (mean feature similarity) and IoU@k (mean spatial overlap) over top-k retrievals:

$$\text{Sim@k} = \frac{1}{k}\sum_{i=1}^{k} \text{sim}(r_{\phi_q}, r_{\phi_{res,i}}),\ \text{IoU@k} = \frac{1}{k}\sum_{i=1}^{k} \text{IoU}(r_{\phi_q}, r_{\phi_{res,i}}),$$

where $r_{\phi_{res,i}}$ is the i-th retrieved region. The mean Time Per Query (mTPQ) is also included in the results. In the region retrieval task, we compare URICA with slide-level, random sampling, and adjacent patch matching baselines. A total of 18,269 randomized queries per site are generated near image trisections, each at least 448 × 448 pixels and rotated within $[-180°, 180°]$, forming a TCGA region retrieval benchmark.

F.2 Extended Quantitative Results

Tab. 2 shows the total results of the WSI slide retrieval task, while Tab. 3 and Tab. 4 show the experiment results of the WSI region retrieval task. For brevity, tissue sites are denoted by abbreviated names: Pul. (Pulmonary), Uri. (Urinary), Mel. (Melanotic), Bra. (Brain), GI. (Gastrointestinal), Liv./PB. (Liver/Pancreas-Biliary), Gyn. (Gynecologic), Pro./Tes. (Prostate/Testis), Hem. (Hematopoiesis), and End. (Endocrine).

Table 2: Slide retrieval experiment results of different methods on different subtypes of TCGA data.

WSI Type	Slide Num	mMV@5					WSI Type	Slide Num	mMV@5				
		Yottixel	SISH	RetCCL	HSHR	**URICA**			Yottixel	SISH	RetCCL	HSHR	**URICA**
Pul.	3395	70.73	68.36	84.27	78.45	**99.26**	**Liv./PB.**	1446	89.05	86.27	89.35	91.97	**95.78**
LUAD	1608	76.34	67.30	78.10	80.11	**99.31**	CHOL	110	34.29	48.08	39.22	37.14	**85.45**
LUSC	1612	69.02	69.62	90.28	77.47	**99.93**	LIHC	870	96.08	93.31	94.97	**96.77**	95.63
MESO	175	35.05	66.47	83.91	72.25	**92.57**	PAAD	466	88.29	81.74	90.83	95.44	**98.49**
Uri.	4198	90.33	88.09	93.67	92.65	**98.80**	**Gyn.**	3610	88.55	85.34	87.14	90.42	**98.95**
BLCA	926	95.97	96.18	**98.37**	96.72	96.00	UCEC	1371	93.34	84.67	81.86	91.65	**99.56**
KIRC	2173	96.24	93.87	93.75	96.29	**99.90**	CESC	604	67.66	69.88	78.32	78.87	**96.68**
KICH	326	84.97	89.85	91.78	83.69	**98.77**	UCS	154	38.31	60.13	68.42	59.48	**96.10**
KIRP	773	69.25	61.45	88.80	81.33	**99.09**	OV	1481	97.93	94.91	93.43	97.22	**99.59**
Mel.	1100	96.61	95.12	93.87	**97.52**	97.36	**Pro./Tes.**	1585	97.38	97.44	98.16	**99.10**	97.92
UVM	150	80.00	79.17	88.41	83.45	**96.67**	TGCT	413	94.26	97.25	98.06	97.99	**99.51**
SKCM	950	99.16	97.57	94.68	**99.68**	97.47	PRAD	1172	98.45	97.50	98.18	**99.48**	97.35
Bra.	3625	93.38	91.60	85.98	93.74	**99.81**	**Hem.**	421	90.38	90.51	90.71	96.40	**99.04**
GBM	2053	93.81	88.86	85.78	95.31	**99.76**	DLBC	103	68.32	79.00	77.97	89.11	**96.11**
LGG	1572	92.83	95.16	86.30	91.70	**99.87**	THYM	318	97.43	94.19	96.77	98.73	**100.0**
GI.	3565	66.97	57.04	54.57	69.22	**97.90**	**End.**	1866	92.49	89.80	94.31	95.19	**98.98**
COAD	1442	87.27	62.28	55.55	76.72	**99.10**	ACC	323	87.85	88.13	81.32	86.96	**98.14**
ESCA	396	72.12	90.26	67.05	69.54	**98.48**	PCPG	385	85.30	86.32	88.89	92.79	**96.88**
READ	530	14.32	16.18	37.27	22.39	**95.28**	THCA	1158	96.49	91.56	98.33	98.49	**99.91**
STAD	1197	67.28	57.64	73.42	81.05	**97.41**	-	-	-	-	-	-	-

Tab.2 shows the mMV@5 results across various TCGA cancer subtypes, demonstrating URICA's superior performance in slide retrieval. URICA outperforms existing approaches (Yottixel, SISH, RetCCL, and HSHR) in most categories, achieving over 95% accuracy across the major organ systems. Notably, it excels in challenging subtypes, such as READ and rare CHOL cases, with performance improvements of up to 88.26% and 37.37%, respectively. URICA also maintains high scalability, delivering outstanding results on both large cohorts (e.g., 99.26% in 3,395 pulmonary slides) and smaller datasets. These results confirm that URICA, leveraging foundation models, achieves state-of-the-art slide retrieval performance on the TCGA dataset, validating its robustness and adaptability across diverse cancer types.

Table 3: Region Retrieval Experiment Results of TCGA with top-1 Retrieval Results.

WSI Site	Query	Slide Method		Sample Method		Adjacent Method		URICA(ours)		WSI Site	Query	Slide Method		Sample Method		Adjacent Method		URICA(ours)	
		mSim@1	mIoU@1	mSim@1	mIoU@1	mSim@1	mIoU@1	mSim@1	mIoU@1			mSim@1	mIoU@1	mSim@1	mIoU@1	mSim@1	mIoU@1	mSim@1	mIoU@1
Pul.	2763	0.7330	0.1981	0.8964	0.4760	0.6502	0.1258	**0.9615**	**0.7108**	**Liv./PB.**	886	0.7142	0.1893	**0.8953**	0.4458	0.5529	0.0915	0.8656	**0.6179**
LUAD	1320	0.7382	0.2002	0.8996	0.4738	0.6700	0.1365	**0.9489**	**0.6302**	CHOL	55	0.6786	0.1532	0.8914	0.4439	0.4646	0.0742	**0.9479**	**0.6977**
LUSC	1335	0.7316	0.1975	0.8937	0.4823	0.6455	0.1217	**0.9615**	**0.7272**	LIHC	520	0.7244	0.1858	**0.9025**	0.4403	0.5422	0.0856	0.8081	**0.5514**
MESO	108	0.7047	0.1885	0.8894	0.4259	0.5371	0.0840	**0.8907**	**0.5422**	PAAD	311	0.7035	0.2015	**0.8840**	0.4554	0.5865	0.1044	0.8819	**0.6446**
Uri.	3209	0.7337	0.1962	0.8979	0.4733	0.6189	0.1154	**0.9409**	**0.7261**	**Gyn.**	2562	0.7324	0.1935	**0.8969**	0.4615	0.5926	0.1025	0.8318	**0.6489**
BLCA	588	0.6944	0.1791	**0.8839**	0.4303	0.5358	0.0815	0.8476	**0.6482**	UCEC	811	0.7223	0.1803	**0.8939**	0.4335	0.5324	0.0767	0.8032	**0.6444**
KIRC	1805	0.7418	0.2008	0.9003	0.4879	0.6352	0.1221	**0.9666**	**0.7461**	CESC	358	0.6827	0.1796	**0.8854**	0.4365	0.4941	0.0743	0.6563	**0.5111**
KICH	265	0.7542	0.2019	0.9020	0.4780	0.6630	0.1385	**0.9715**	**0.6551**	UCS	95	0.7343	0.1806	0.8911	0.4302	0.5555	0.0653	**0.9645**	**0.7104**
KIRP	551	0.7396	0.1964	0.9029	0.4689	0.6329	0.1182	**0.9712**	**0.7768**	OV	1298	0.7522	0.2065	0.9025	0.4881	0.6601	0.1292	**0.9652**	**0.7363**
Mel.	645	0.6848	0.1807	**0.8839**	0.4276	0.5333	0.0806	0.7414	**0.5543**	**Pro./Tes.**	1126	0.6866	0.1932	**0.8830**	0.4562	0.5726	0.1067	0.8701	**0.6573**
UVM	97	0.6857	0.1833	0.8878	0.3889	0.5056	0.0603	**0.9648**	**0.7394**	TGCT	292	0.6972	0.1966	**0.8836**	0.4516	0.5614	0.1049	0.7359	**0.5424**
SKCM	548	0.6829	0.1803	**0.8832**	0.4345	0.5382	0.0842	0.7134	**0.5312**	PRAD	834	0.6829	0.1920	0.8828	0.4578	0.5765	0.1074	**0.9184**	**0.6986**
Bra.	2939	0.7418	0.1918	0.9055	0.4606	0.6177	0.1072	**0.9473**	**0.7232**	**Hem.**	286	0.7265	0.1907	0.8957	0.4544	0.5547	0.0927	**0.8994**	**0.6948**
GBM	1739	0.7297	0.1916	0.8995	0.4611	0.6101	0.1081	**0.9243**	**0.6997**	DLBC	77	0.7133	0.1810	0.8998	0.4644	0.5923	0.0986	**0.9634**	**0.7117**
LGG	1200	0.7594	0.1920	0.9142	0.4599	0.6288	0.1058	**0.9763**	**0.7528**	THYM	209	0.7314	0.1943	**0.8941**	0.4507	0.5408	0.0905	0.8807	**0.6899**
GI.	2690	0.7056	0.1979	0.8880	0.4724	0.5968	0.1185	**0.8901**	**0.6657**	**End.**	1163	0.7151	0.1806	**0.8941**	0.4435	0.5689	0.0918	0.8931	**0.6396**
COAD	1140	0.7162	0.2033	0.8895	0.4775	0.5973	0.1199	**0.8953**	**0.6655**	ACC	229	0.7495	0.1879	0.8969	0.4744	0.6330	0.0948	**0.9637**	**0.7611**
ESCA	215	0.6805	0.1848	**0.8808**	0.4369	0.4849	0.0813	0.6362	**0.5595**	PCPG	242	0.7203	0.1732	0.8998	0.4202	0.5780	0.0907	**0.9597**	**0.7134**
READ	389	0.7052	0.1947	0.8880	0.4693	0.5926	0.1171	**0.9702**	**0.6922**	THCA	692	0.7019	0.1807	0.8912	0.4414	0.5445	0.0911	**0.9056**	**0.7029**
STAD	946	0.6987	0.1958	**0.8877**	0.4757	0.6233	0.1259	0.8740	**0.6659**	-	-	-	-	-	-	-	-	-	-

Tab.3 and Tab.4 show the region retrieval results in the TCGA dataset with top-1 and top-3 retrieval results. It is worth noting that the adjacent method's effect is close to that of the slide method when the length of the adjoining column is only 1, but it does not support the accuracy of the result for long columns. While the slide method and URICA have almost the same performance under different column lengths. URICA is far superior to the slide method and the adjacent method in most cases. These results show that URICA shows state-of-the-art results in other areas.

As shown in Tab. 3, we evaluate the region retrieval performance of URICA and three baseline methods—Slide Method, Sample Method, and Adjacent Method—on the TCGA dataset using top-1 retrieval. URICA consistently achieves the highest mSim@1 and mIoU@1 scores across nearly all tissue types, demonstrating its superior capability in capturing both semantic and structural consistency between the query region and the retrieved results. Although the Adjacent Method shows comparable performance to the Slide Method when the adjacent column length equals one, its performance rapidly degrades with increasing column length due to the lack of global contextual information. By

Table 4: Region Retrieval Experiment Results of TCGA with top-3 Retrieval Results.

WSI Site	Query	Slide Method		Sample Method		Adjacent Method		URICA(ours)		WSI Site	Query	Slide Method		Sample Method		Adjacent Method		URICA(ours)	
		mSim@3	mIoU@3	mSim@3	mIoU@3	mSim@3	mIoU@3	mSim@3	mIoU@3			mSim@3	mIoU@3	mSim@3	mIoU@3	mSim@3	mIoU@3	mSim@3	mIoU@3
Pul.	2763	0.7253	0.1909	0.8744	0.3867	0.2774	0.0436	**0.9549**	**0.6819**	**Liv./PB.**	886	0.7007	0.1751	**0.8720**	0.3457	0.2328	0.0315	0.8582	**0.5868**
LUAD	1320	0.7307	0.1933	0.8789	0.3881	0.2861	0.0478	**0.9559**	**0.6698**	CHOL	55	0.6593	0.1374	0.8670	0.3377	0.1801	0.0253	**0.9420**	**0.6609**
LUSC	1335	0.7244	0.1904	0.8705	0.3890	0.2736	0.0419	**0.9569**	**0.6991**	LIHC	520	0.7095	0.1708	**0.8803**	0.3361	0.2306	0.0296	0.8047	**0.5201**
MESO	108	0.6900	0.1758	0.8672	0.3412	0.2487	0.0294	**0.9005**	**0.5756**	PAAD	311	0.6935	0.1891	0.8590	0.3631	0.2458	0.0359	**0.8697**	**0.6166**
Uri.	3209	0.7273	0.1962	0.8758	0.3791	0.2597	0.0399	**0.9344**	**0.6856**	**Gyn.**	2562	0.7216	0.1824	0.8753	0.3656	0.2468	0.0353	**0.8079**	**0.6029**
BLCA	588	0.6799	0.1693	**0.8587**	0.3262	0.2210	0.0280	0.8429	**0.6195**	UCEC	811	0.7093	0.1674	**0.8717**	0.3277	0.2136	0.0262	0.7520	**0.5825**
KIRC	1805	0.7384	0.1969	0.8786	0.3951	0.2668	0.0423	**0.9599**	**0.7073**	CESC	358	0.6600	0.1570	**0.8604**	0.3299	0.2142	0.0257	0.6514	**0.4689**
KICH	265	0.7518	0.1994	0.8803	0.3865	0.2818	0.0477	**0.9522**	**0.6049**	UCS	95	0.7225	0.1686	0.8678	0.3439	0.2364	0.0224	**0.9634**	**0.6548**
KIRP	551	0.7295	0.1875	0.8828	0.3796	0.2672	0.0409	**0.9654**	**0.7118**	OV	1298	0.7461	0.1998	0.8821	0.4007	0.2772	0.0446	**0.9595**	**0.7075**
Mel.	645	0.6752	0.1691	**0.8590**	0.3245	0.2313	0.0281	0.7346	**0.5190**	**Pro./Tes.**	1126	0.6800	0.1857	0.8584	0.3663	0.2326	0.0368	**0.8654**	**0.6212**
UVM	97	0.6840	0.1718	0.8622	0.2907	0.2179	0.0211	**0.9626**	**0.7291**	TGCT	292	0.6903	0.1881	**0.8592**	0.3607	0.2303	0.0364	0.7334	**0.5086**
SKCM	548	0.6736	0.1686	**0.8584**	0.3304	0.2337	0.0293	0.7061	**0.4927**	PRAD	834	0.6764	0.1849	0.8581	0.3682	0.2334	0.0369	**0.9130**	**0.6617**
Bra.	2939	0.7331	0.1827	0.8861	0.3632	0.2581	0.0370	**0.9288**	**0.6805**	**Hem.**	286	0.7178	0.1830	0.8734	0.3511	0.2478	0.0323	**0.8937**	**0.6545**
GBM	1739	0.7207	0.1822	0.8790	0.3654	0.2542	0.0374	**0.8944**	**0.6508**	DLBC	77	0.7030	0.1747	0.8795	0.3719	0.2699	0.0347	**0.9594**	**0.6719**
LGG	1200	0.7112	0.1575	0.8964	0.3602	0.2636	0.0364	**0.9723**	**0.7181**	THYM	209	0.7232	0.1860	0.8712	0.3434	0.2397	0.0315	**0.8745**	**0.6494**
GI.	2690	0.6974	0.1900	0.8630	0.3792	0.2442	0.0408	**0.8716**	**0.6295**	**End.**	1163	0.7042	0.1705	0.8719	0.3431	0.2435	0.0316	**0.9064**	**0.6739**
COAD	1140	0.7088	0.1944	0.8648	0.3858	0.2426	0.0412	**0.8675**	**0.6281**	ACC	229	0.7445	0.1828	0.8738	0.3740	0.2786	0.0329	**0.9060**	**0.6986**
ESCA	215	**0.6622**	0.1681	0.8527	0.3339	0.2002	0.0281	0.6078	**0.5283**	PCPG	242	0.7076	0.1613	0.8794	0.3276	0.2560	0.0314	**0.9467**	**0.6621**
READ	389	0.6929	0.1855	0.8653	0.3727	0.2394	0.0404	**0.9654**	**0.6620**	THCA	692	0.6897	0.1696	0.8686	0.3383	0.2276	0.0312	**0.8981**	**0.6621**
STAD	946	0.6934	0.1914	0.8623	0.3841	0.2581	0.0434	**0.8693**	**0.6297**	-	-	-	-	-	-	-	-	-	-

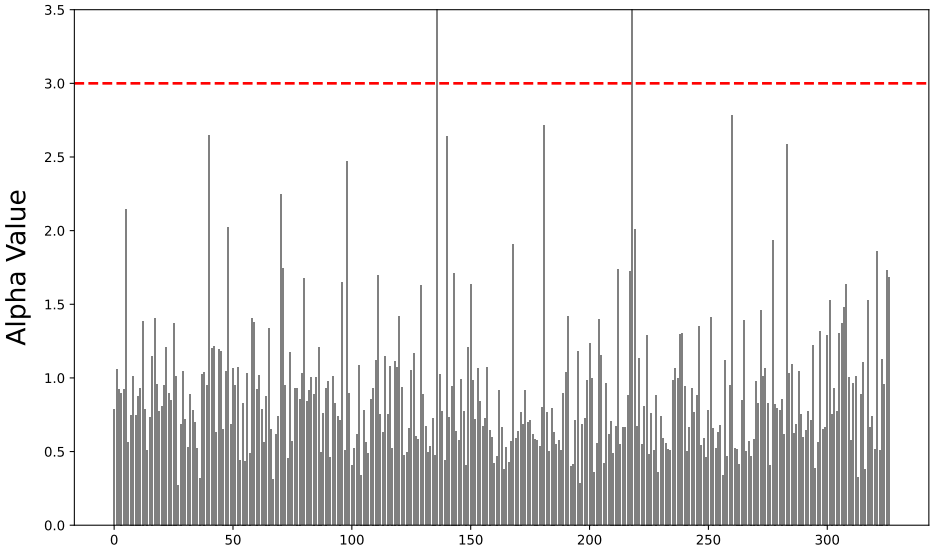


Figure 2: The alpha observation result in hematopoiesis WSIs of TCGA-DLBC and TCGA-THYM.

contrast, URICA maintains stable performance across various tissue domains, such as Pulmonary, Brain, and Gastrointestinal, exhibiting strong generalization and robustness to intra-class variability. These results validate that URICA effectively balances local adjacency and global semantic alignment, outperforming traditional slide-level and adjacency-based retrieval strategies.

Similarly, Tab. 4 presents the top-3 retrieval results, which further confirm the effectiveness of URICA. Even when expanding the retrieval scope to the top-3 candidates, URICA continues to deliver the best mSim@3 and mIoU@3 scores, indicating that its retrieved regions are not only semantically consistent but also spatially coherent with the query. The Slide and Adjacent Methods, in contrast, experience a notable decline in both similarity and overlap metrics, reflecting their limited discriminative power when dealing with subtle inter-region variations. Particularly in tissue types with complex morphology (e.g., Brain, Endocrine, and Gynecologic sites), URICA maintains high performance stability, revealing its ability to capture histological relevance beyond local pixel similarity. Overall, these results highlight that URICA achieves state-of-the-art region retrieval performance in both strict (top-1) and relaxed (top-3) settings.

To complement the main paper, we further report per-site behavior directly in Tab. 3 and Tab. 4: URICA provides consistent gains on most high-variance sites (Bra., End., Liv./PB.) and remains competitive in edge cases such as ESCA/CESC where inter-class morphology is highly overlapping.

F.3 Alpha Estimation Details

Based on Hyp. 1, we estimate α by sampling 100 sets of $(224, 224)$ patch pairs from 421 hematopoietic WSIs (TCGA-DLBC and TCGA-THYM) across magnifications, using the UNI encoder $E^*_{224}(\cdot)$. Each set consists of patches $(r_{\phi_a}, r_{\phi_b}, r_{\phi_c}, r_{\phi_d})_i$ with fixed angular alignment ($\theta = 0$) and controlled displacements: $x_a - x_c = x_b - x_d \in [-112, 112]$, $y_a - y_c = y_b - y_d \in [-112, 112]$. For each set, we compute the cosine similarities $\rho_{ac_i} = |sim(a, c)|$ and $\rho_{bd_i} = |sim(b, d)|$ and derive α via:

$$\alpha = \max(\{\alpha_i\}), \quad \alpha_i \quad = \begin{cases} \frac{\ln(1-\rho_{ac_i}+\epsilon)-\ln(1-\rho_{bd_i}+\epsilon)}{\ln(\delta+\epsilon)} & \text{if } \rho_{ac_i} < \rho_{bd_i} \ , \\ \frac{\ln(\rho_{ac_i}+\epsilon)-\ln(\rho_{bd_i}+\epsilon)}{\ln(\delta+\epsilon)} & \text{if } \rho_{ac_i} > \rho_{bd_i} \ , \\ 0 & \text{otherwise} \ , \end{cases}$$

where $\delta = (1 - (x_a - x_c)/224)(1 - (y_a - y_c)/224)$ is the degree of coincidence based on Hyp. 1, and $\epsilon > 0$ is an infinitesimal constant. We visualize the α values across the sampled WSIs. The results suggest that semantic similarity decays non-linearly with spatial displacement, supporting the need for a calibrated tessellation stride.

The empirical α_i distribution is right-skewed: most values concentrate in a stable middle range while a small fraction of hard pairs produces larger exponents. We therefore use a robust summary (median and interquartile behavior) rather than only extreme values, and set the default coefficient to $\alpha = 3$ as a stable operating point for cross-site experiments.

Sensitivity to stride t. We observe the same trend as predicted by the analysis section: when t is too small, computation grows with limited accuracy gain; when t is too large, similarity bounds become loose and mIoU decreases. Around $t \approx 60$, the trade-off between quality and cost is consistently strong under $\alpha = 3$, so this value is used as the default in the reported benchmark.

F.4 Additional Visualizations and Failure Cases

Fig. 3 presents a comparative visualization of three retrieval paradigms: adjacent-method retrieval, slide-thumbnail retrieval, and URICA region retrieval. The URICA method exhibits markedly higher regional consistency with the query patches, both in morphology and semantics. In contrast, the adjacent method relies primarily on local patch adjacency; it therefore focuses on fine-grained textures but fails to reconstruct coherent regions when the retrieved patches are sparsely distributed, leading to incomplete structural continuity. The slide-thumbnail retrieval can capture globally similar tissue contexts at the slide level, yet lacks the precision to localize semantically and structurally matched sub-regions within those slides.

By bridging these two extremes, URICA achieves region-level retrieval that aligns well with both local structural features and higher-order semantic patterns. This unified representation enables more faithful region reconstruction and provides a stronger foundation for downstream computational-pathology tasks such as spatial analysis, phenotype clustering, and morphological comparison.

We additionally observe two recurring failure patterns: (1) highly repetitive stromal textures causing false-positive high similarity across different subtypes, and (2) very small lesion regions where anchor coverage is insufficient after rotation. These cases mainly affect IoU while preserving moderate semantic similarity, suggesting future improvements should combine adaptive anchor density with uncertainty-aware re-ranking.

F.5 Human Study Protocol

To validate whether the automated retrieval metrics align with expert clinical judgment, we conduct a structured human evaluation study with board-certified pathologists.

Study design. For each of the 10 anatomical sites, we randomly sample evaluation cases and present the pathologist with the query region alongside the top-5 retrieved regions produced by URICA. The study is conducted through a blinded interface: pathologists are unaware of the retrieval method and ranking order. For each case, the pathologist (i) assigns a similarity score (0–10 scale) to every retrieved region, and (ii) provides a preferred ranking of the five candidates.

Evaluation metrics. We define three complementary metrics to assess different aspects of retrieval quality:

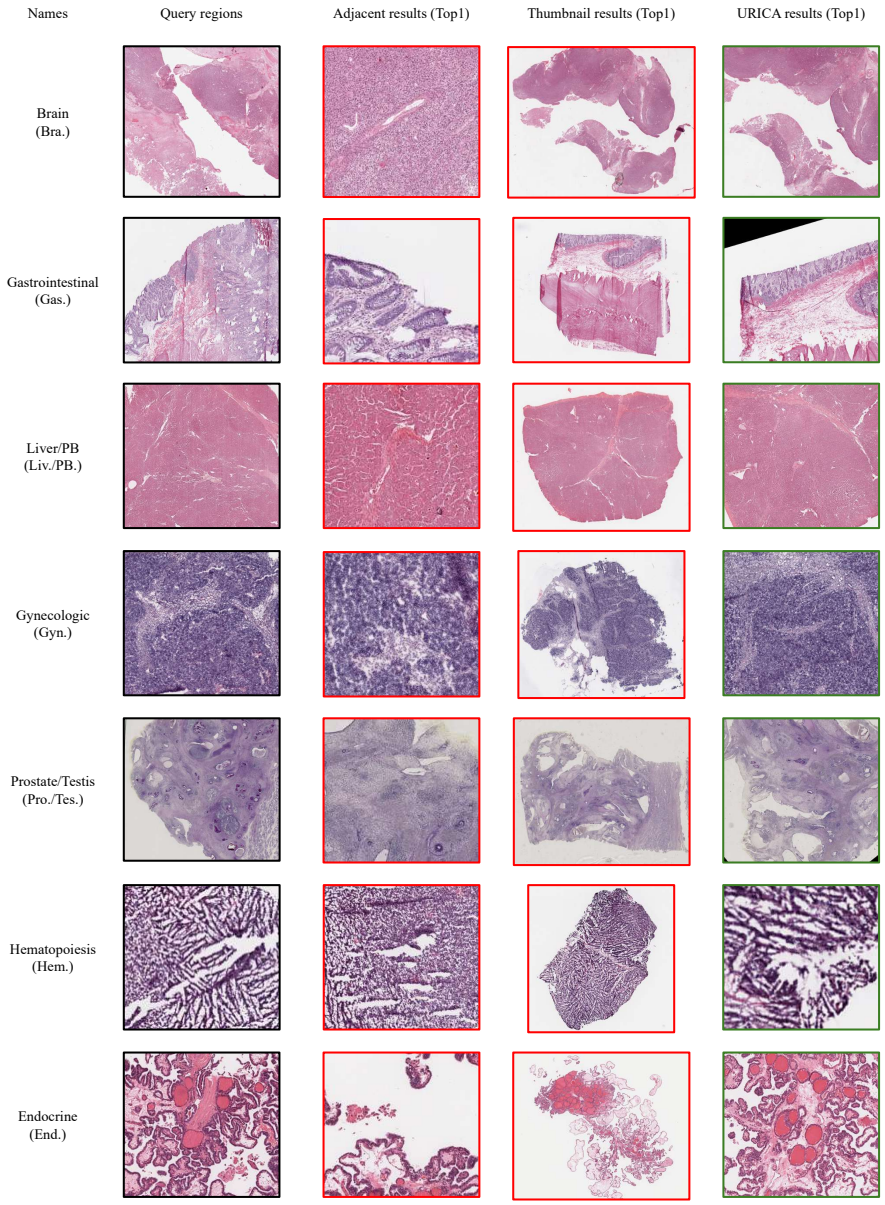


Figure 3: Qualitative visualization of region retrieval results across different methods.

- **Order Quality** evaluates whether the model ranking agrees with the pathologist ranking. Given a pathologist permutation of the top-5, we compute the inversion ratio inv_ratio = inv_count/10 (where 10 is the maximum number of inversions in a 5-element permutation) and define order_quality = $1 - \text{inv\_ratio} \in [0, 1]$, where 1 indicates perfect agreement.
- **Mean Similarity** is the average pathologist-assigned score across the five retrieved regions per case, reflecting overall candidate quality.
- **Intra-Top-5 Consistency** measures the uniformity of scores within the top-5 set. We compute the mean pairwise difference among the five scores, normalize it, and define consistency as $1 -$ normalized_mean_pairwise_diff $\in [0, 1]$. Higher values indicate that the top-5 candidates are of uniformly high (or low) quality, rather than a mix of good and poor results.

Results. Tab. 5 summarizes the per-site and overall results. Across all 10 sites, URICA achieves an overall mean similarity of **7.09** (out of 10), an intra-top-5 consistency of **0.91**, and an order quality of **0.57**.

The high mean similarity score indicates that pathologists generally regard the retrieved regions as relevant to the query, corroborating the automated mSim metric. The consistently high intra-top-5

Table 5: Per-site human evaluation summary. Mean Similarity is the average pathologist-assigned similarity score (scale 0–10). Intra-Top-5 Consistency measures how uniform the top-5 scores are (1 = identical). Order Quality measures agreement between model and pathologist ranking (1 = perfect).

Site	Mean Similarity	Intra-Top-5 Consistency	Order Quality
Pulmonary	7.1300	0.9111	0.5400
Urinary	7.0100	0.9078	0.4850
Melanocytic	5.9600	0.8800	0.4900
Brain	7.5000	0.9344	0.6550
Gastrointestinal	7.7400	0.9367	0.6200
Liver/PB	6.5300	0.8567	0.6450
Gynecologic	7.5900	0.9267	0.5550
Prostate/Testis	6.9400	0.9044	0.4500
Hematopoietic	7.2800	0.9067	0.6000
Endocrine	7.2000	0.9411	0.6550
Overall	**7.0880**	**0.9106**	**0.5695**

consistency (> 0.85 for all sites) demonstrates that URICA does not merely retrieve a single good match surrounded by poor candidates; instead, the entire top-5 set maintains comparable quality. The order quality, while moderate in absolute value, reflects the inherent subjectivity in fine-grained ranking among already-similar candidates.

Fig. 4 visualizes the case-level distribution of order quality per site. Sites such as Brain and Endocrine exhibit higher median order quality (0.655), while Prostate/Testis and Urinary show more variability, suggesting these sites contain morphologically ambiguous cases where both model and human rankings are less deterministic. The interquartile spread is generally narrow, indicating stable ranking performance across cases within each site.

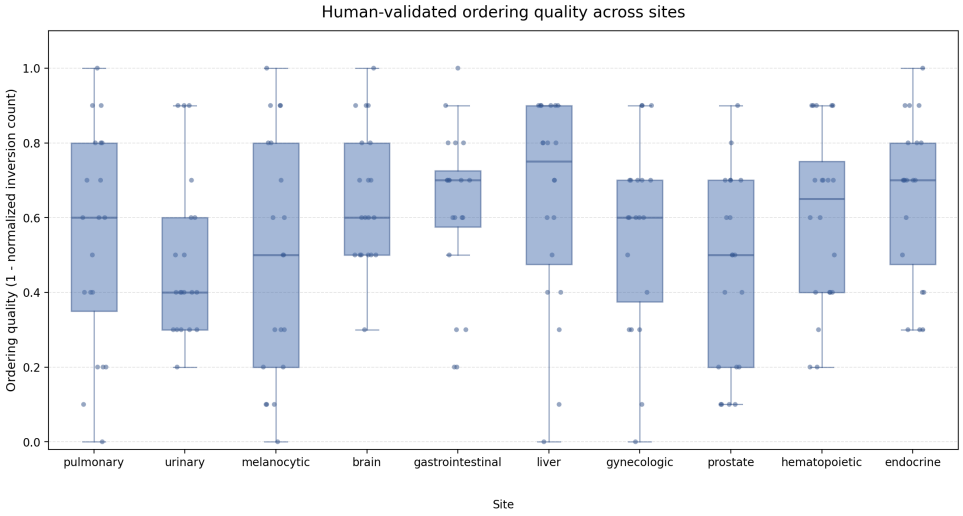


Figure 4: Per-site distribution of order quality scores. Each box summarizes the case-level order quality within a site; overlaid dots denote individual cases. A higher median indicates better agreement between the model ranking and the pathologist ranking.

Fig. 5 provides a complementary perspective by decoupling ranking from set quality. High blue bars (mean similarity) combined with high orange bars (consistency) confirm that the top-5 candidates are collectively strong and uniform. Notably, Gastrointestinal and Brain achieve both the highest mean similarity (7.74 and 7.50) and the highest consistency (0.937 and 0.934), whereas Melanocytic shows relatively lower similarity (5.96) and consistency (0.880), likely due to the morphological diversity within melanocytic lesions and limited database coverage.

Summary. The two figures provide complementary evidence: Fig. 4 addresses "*is the ranking correct?*" (ordering ability), while Fig. 5 addresses "*is the candidate set itself good?*" (independent of ordering). Together, they allow us to distinguish ranking errors from candidate quality issues and from

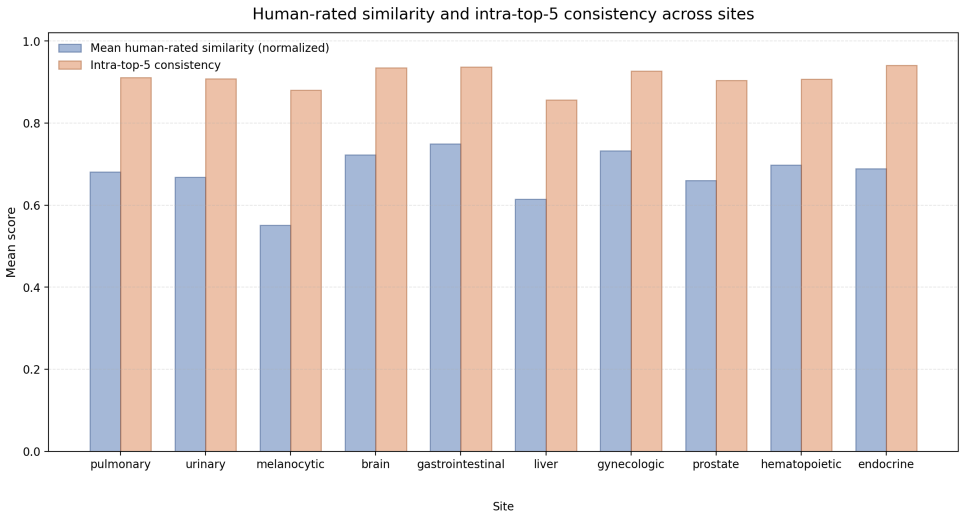


Figure 5: Per-site normalized mean similarity (blue) and intra-top-5 consistency (orange). High blue bars indicate that the retrieved candidates are rated as highly similar by pathologists; high orange bars indicate that the top-5 scores are internally consistent.

database coverage limitations. The human evaluation confirms the same performance trends observed in automated metrics and validates that URICA retrieval results are clinically meaningful.