URICA: A Uniformity Region Affine Identifier Capture Algorithm for Arbitrary Region Retrieval in Pathology Images

Ri Su[1] Zhao Chen[1,*] Caleb Chen Cao[2,*] Lei Chen[1,2]
[1] HKUST(GZ), China, [2] HKUST, Hong Kong SAR, China
rsu704@connect.hkust-gz.edu.cn, chenzhao@hkust-gz.edu.cn, caochen.hkust@gmail.com, leichen@cse.ust.hk

Abstract

Whole slide image (WSI) region retrieval remains an open challenge in computational pathology, as existing methods struggle to represent and preserve information of all possible regions. Current approaches that rely on fixed-size patches or slide-level retrieval are misaligned with real clinical workflows, where pathologists often examine WSI regions of arbitrary orientations and sizes rather than predefined patches or slides. In this work, we redefine WSI retrieval as a semantically optimal matching problem between arbitrary regions under spatial transformations, which necessitates a region-level representation that maintains semantic consistency. To fulfill this requirement, we introduce semantic tessellation, which organizes patch units into flexible, geometry-aware region descriptors. Building on this representation, we develop the affine identifier, a semantic signature that enables rotation- and scale-consistent region matching. We further derive theoretical bounds between the tessellation-derived descriptors and the ideal pixel-level semantic mask objective, showing that they reliably approximate mask-based region similarity. Together, these components form URICA, a theoretically grounded algorithm for robust WSI region retrieval. Experiments on large public datasets demonstrate that URICA achieves strong and consistent performance across diverse WSI retrieval tasks[1].

1. Introduction

Whole slide image (WSI) [22] is a gigapixel scan of histological specimens that enables computational examination of tissue architecture, forming the foundation of modern digital pathology pipelines [20]. The large size and heterogeneous tissue organization of WSI make its regions highly variable in orientation and size, requiring representations that remain consistent under such spatial variations. Yet current retrieval methods do not meet these region-level and transformation-consistent requirements. Patch-based approaches [14, 37] estimate similarity from isolated patches and therefore lack the spatial context required to reconstruct coherent regions. Slide-level representations [19, 38] collapse fine-grained morphology into global representations. As a result, neither of these methods can retrieve region-specific patterns, which is an essential capability for identifying clinically relevant conditions such as mucinous or ductal carcinoma in situ [11]. This limitation arises from a more fundamental difficulty: WSIs contain no predefined objects, and region boundaries, orientations, and scales vary freely, making it hard to build representations that preserve region-level semantics under varying conditions.

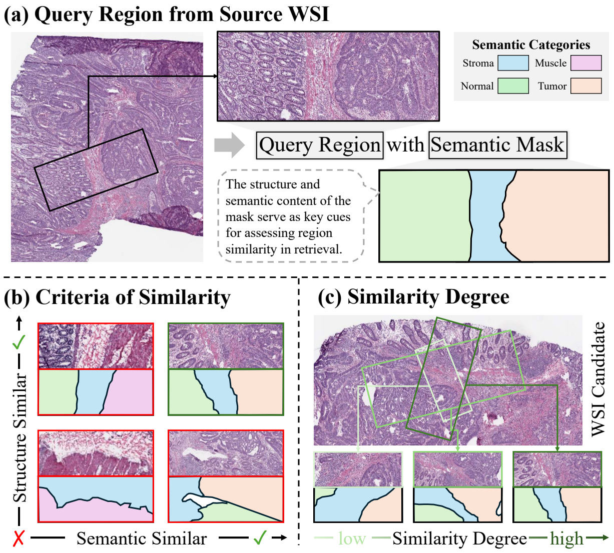

Figure 1. A region retrieval example from colorectal tissue shows a boundary between normal and tumor areas. WSI region retrieval aims to (a) select the query region with its semantic mask, (b) find regions exhibiting both semantic and structural similarity, and (c) identify the most similar candidate region.

The first challenge lies in representing arbitrary regions. Segmentation-based methods [16] represent regions with pixel-level masks of histological components, often guided

---

*Corresponding authors.

[1] Code is available at https://github.com/HKUSTMDI/URICA-CVPR26

by weakly supervised [26] or SAM-based models [15]. While such masks can express region boundaries to some extent, they cannot enumerate or store region representations across all possible scales and rotations—making them unsuitable for various transformation scenarios. By contrast, self-supervised frameworks [24] such as UNI [7] and PathDino [1] provide robust semantic features and basic units via large-scale pretraining. However, they offer patch-level semantics but no explicit way to compose these units into a consistent representation of arbitrary regions.

Another challenge is maintaining spatial and semantic consistency during retrieval, as existing systems [5, 34] cannot reliably match regions when rotation or scale varies. Fig. 1 illustrates that, despite variations in rotation and scaling, retrieved regions with identical normal, stroma, and tumor semantics (dark green box) exhibit both structural and semantic similarity to the query image, whereas regions with distinct patterns (red boxes) fail to represent structure and semantic correspondence simultaneously. Although rotation-aware object detection methods [8, 12, 40] provide partial solutions, they are ineffective in histopathology due to the lack of well-defined objects. Taken together, these challenges reveal a fundamental gap between current retrieval paradigms and the fine-grained region-level demands of computational pathology, underscoring the need for robust region representation and management—an ability still largely absent in existing WSI analysis pipelines.

To address these challenges, we first formulate WSI region retrieval as a semantically optimal matching problem under arbitrary spatial transformations, reflecting the need to maintain spatial and semantic consistency across rotation and scale variations. Meeting this objective requires a region-level representation that can express arbitrary regions and remain stable under transformation. To satisfy this representation requirement, we introduce a semantic tessellation that composes foundation-model patch features into flexible descriptors capable of capturing regions while representing their underlying semantic structure. Building on the tessellation, we derive a transformation-invariant region signature that captures consistent semantic relationships under rotation and magnification, enabling robust region matching. We further show that the tessellation representation closely approximates ideal pixel-level mask matching, providing a principled bridge between practical descriptors and the ideal objective. These components together form our approach, **URICA** (Uniformity Region Affine Identifier Capture Algorithm).

The key contributions of this work are:

- First, we introduce tessellation, a novel structural concept that partitions WSIs into geometrically consistent units, enabling efficient and accurate region retrieval.
- Second, we design URICA for semantic region retrieval under arbitrary rotations and scaling, leveraging affine identifier consistency and complementary techniques such as the anchor selection method.
- Third, we provide a theoretical analysis demonstrating that URICA approximates pixel-level semantic mask retrieval based on tessellation structure, and give efficiency improvements ensuring optimal applications.

Experiments across diverse conditions on 29 TCGA cancer subtypes across 10 anatomical sites show that URICA achieves state-of-the-art performance on both slide- and region-level tasks. To our knowledge, this is the first approach to support the retrieval of arbitrary WSI regions.

2. Related Work

Pathology Image Representation. Computational pathology is challenged by gigapixel WSIs and limited annotations. Self-supervised ViTs, including DINO-based models [6] and CTransPath [33], learn morphological features via contrastive objectives, while REMEDIS [2] combines supervised transfer with contrastive adaptation. Foundation models such as UNI [7] and PathDino [1] improve generalization and rotational robustness through large-scale pretraining, with further enhancements from feature re-embedding [27] and morphological prototyping [25]. Graph-based methods [4, 9] capture tissue relationships, and weak supervision (e.g., CLAM [21], dual-stream MIL [17]) integrate multi-scale semantics for slide-level aggregation. Yet they remain patch- or slide-centric and lack explicit consistency under transformations.

WSI Retrieval. The growth of digital pathology has spurred scalable WSI retrieval for diagnostic and research use. Yottixel [14] introduces a Bunch of Barcodes for efficient indexing, while DRA-Net [37] uses attention to highlight diagnostic regions. SISH [5], RetCCL [34], and HSHR [19] adopt contrastive or hybrid self-supervised learning to improve robustness and precision. For content-based retrieval, Zheng *et al.* [36] developed a size-scalable system, and hierarchical graph models [38, 39] preserve spatial relations via location-aware hashing; Hu *et al.* [13] further fuse reports and metadata for explainable retrieval. System advances include in-memory spatial indexing [28], high-performance analytics [29], and hybrid GPU–CPU architectures [31]. Nonetheless, existing systems are largely slide-level and do not ensure semantic or structural alignment under rotated or rescaled regions.

Oriented Object Detection. Rotation-aware methods model orientation for geometric invariance. GIMI [18] incorporates geo-location into contrastive learning, and Steerers [3] learn rotation-equivariant descriptors. ReDet [10], AO2-DETR [8], and SA3Det [32] improve detection via equivariant networks and adaptive proposals, while Fourier-based contour modeling [40] refines orientation sensitivity. However, these methods focus on discrete objects, while WSI tissue regions are continuous and dense.

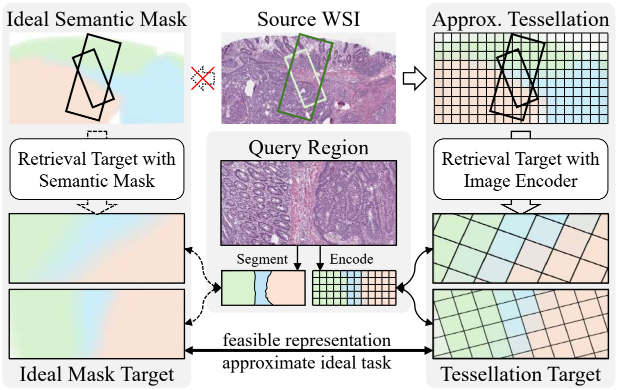

Figure 2. Comparison between the ideal mask–based retrieval and the feasible tessellation–based retrieval. The latter replaces oracle semantic masks with encoder-derived patch embeddings.

3. Preliminaries

WSI region retrieval addresses the challenge of identifying the most similar region under varying rotations and scales. While pixel-level semantic masks are ideal but impractical, encoder-based features offer tractable yet sparse patch-level approximations. To bridge this gap, we define the semantic tessellation, which aligns encoder-based representations with the ideal mask formulation. Both similarity definitions are formalized below and illustrated in Fig. 2.

3.1. Region with Segmentation Mask

A WSI I is acquired through high magnification scanning at size $W \times H$. Given this scanned image, a pathology image $I_l \in \mathbb{R}^{W_l \times H_l \times 3}$ is obtained as $I_l = D(I, l)$ using a down-sampling function D at level l. In particular, the standard formats (e.g., .svs, .tiff) organize $\{I_l | I\}$ into a pyramid structure and fixed indices for faster visualization. A WSI set $\mathcal{I} = \{I\}_{i=1}^{n}$ then consists of n such WSIs.

Given an I_l and an ideal segmentation model $Seg(\cdot)$, an ideal semantic mask $M_l = Seg(I_l) \in \mathbb{R}^{W_l \times H_l \times d}$ is a pixel-level mask corresponding to I_l, 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) in I_l. Based on this, we define a region:

Definition 1 ***(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 a descriptor $\phi = (x, y, w, h, \theta)$ with:*

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

where (x, y) and (w, h) are the center position and size of regions r_ϕ or m_ϕ, respectively, with $\theta \in [-\pi, \pi)$, $0 \leq x - w/2 < x + w/2 \leq W_l$, and $0 \leq y - h/2 < y + h/2 \leq H_l$.

3.2. Region with Tessellation

While ideal region retrieval would match regions with m_ϕ differences, obtaining pixel-level masks relies on perfect segmentation models, which are rarely feasible. In practice, vision foundation models can represent region semantics at the patch level. Here, we redefine the semantic mask and formally define the image encoder.

Definition 2 ***(Semantic Mask)****: Given a pathology image $I_l \in \mathbb{R}^{W_l \times H_l \times 3}$, the semantic mask $M_l \in \mathbb{R}^{W_l \times H_l \times S \times d_s}$ stores semantic information of I_l in semantic space $\mathbb{R}^{d_s}$ over granularity ratio $s \in [0, S]$. Here $M_l(x, y, s) \in \mathbb{R}^{d_s}$ denotes the semantic representation at location (x, y) and granularity s, corresponding to a local descriptor $\phi_s = (x, y, 2s, 2s, \theta)$ with arbitrary θ. A semantic region is denoted by $m^s_\phi \subseteq M_l(s)$ for descriptor $\phi = (x, y, w, h, \theta)$.*

Definition 3 ***(Image Encoder)****: Given an I_l and corresponding M_l, a model $E_{s^*}(r_\phi)$ is an image encoder if $\forall \;\; \phi = (x, y, 2s, 2s, \theta), \;\; s \geq s^*, \; \exists$ a global function $L^s(\cdot) : \mathbb{R}^{2s \times 2s} \to \mathbb{R}^{d_s}$ such that $E_{s^*}(r_\phi) = L^s(m_\phi) = e_\phi \in \mathbb{R}^{d_s}$, where s^* is the minimum semantic granularity ratio of $E_{s^*}(r_\phi)$, and d_s is the size of embedding e_ϕ.*

In Def. 3, the encoder maps a region r_ϕ to its semantic representation m_ϕ at minimum granularity s^*. Existing encoders struggle when $s < s^*$, so we introduce region tessellation to support semantics at arbitrary granularity.

Definition 4 ***(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 $T^t_\phi = \{V_\phi, R_\phi\}$ is an overlapping square tiling composed of sampled semantic region anchors m_{ϕ_g} with a step size t and their adjacency relations:*

$$V_\phi = \{(m_{\phi_g}, x_g, y_g) \mid \phi_g = (x_g, y_g, 2s, 2s, \theta), \; m^s_\phi(x_g, y_g)\},$$
$$R_\phi = \{(m_{\phi_u}, m_{\phi_v}) \mid m_{\phi_u}, m_{\phi_v} \in V_\phi, \; Adj(m_{\phi_u}, m_{\phi_v})\},$$

where $T^t_\phi(x_g, y_g) = m_{\phi_g}$, $(x_g, y_g) = (x + x' \cdot \cos(\theta) - y' \cdot \sin(\theta), \;\; y + x' \cdot \sin(\theta) + y' \cdot \cos(\theta))$ 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 using the Euclidean axes of r_ϕ

From a representational perspective, the region tessellation samples anchor points from m^s_ϕ with step size t, capturing relative spatial relationships within the semantic mask.[2]

4. URICA: Region Retrieval Algorithm

To approximate mask-based region retrieval within the image encoder (Def. 3) and the tessellation-based retrieval

---

[2] The complete definitions and notations are provided in Appendix B

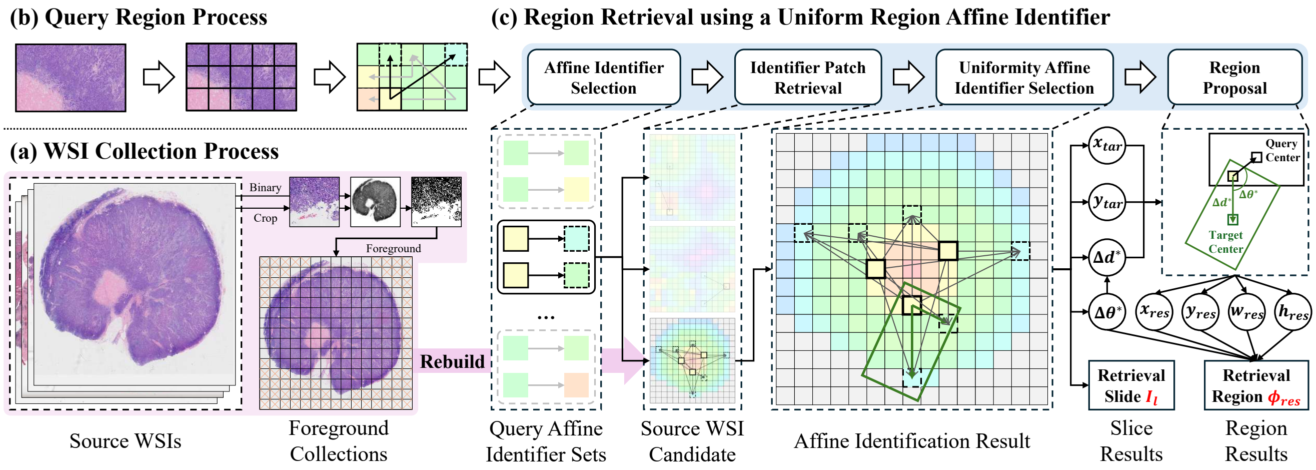

Figure 3. Overview of the URICA pipeline for WSI region retrieval. (a) Source WSIs are preprocessed into a foreground patch vector database. (b) Each query region is decomposed into tessellated patches and assigned spatial affine identifiers. (c) The retrieval process matches affine identifier sets across candidate WSIs and reconstructs the target region ϕ_{res} based on the most uniform affine alignment.

framework (Def. 4), we first analyze properties that are invariant to rotation and scale. For any two corresponding region r_ϕ and $r_{\phi'}$, we have the following property:

Property 1 *Given two regions r_ϕ and $r_{\phi'}$ with $\phi = (x, y, w, h, \theta)$ and $\phi' = (x', y', w \times \Delta d, h \times \Delta d, \theta + \Delta\theta)$, $\forall\ \Delta\theta \in [-\pi, \pi)$ and $\Delta d > 0$, an affine identifier $p(\cdot) = (x - x_0, y - y_0)$ denotes the vector from (x_0, y_0) to (x, y) within the tessellation. For any $p(e) = (x_e, y_e) \in r_\phi$ and corresponding $p(e') = (x_{e'}, y_{e'}) \in r_{\phi'}$, we have:*

$$\sqrt{{x_{e'}}^2 + {y_{e'}}^2} / \sqrt{{x_e}^2 + {y_e}^2} = \Delta d,$$

$$\arccos\left(\frac{x_{e'}}{\sqrt{{x_{e'}}^2 + {y_{e'}}^2}}\right) - \arccos\left(\frac{x_e}{\sqrt{x_e^2 + y_e^2}}\right) = \Delta\theta,$$

where

$$\begin{bmatrix} x_e \\ y_e \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x^* \\ y^* \end{bmatrix}$$

$$\begin{bmatrix} x_{e'} \\ y_{e'} \end{bmatrix} = \Delta d \begin{bmatrix} \cos(\theta + \Delta\theta) & -\sin(\theta + \Delta\theta) \\ \sin(\theta + \Delta\theta) & \cos(\theta + \Delta\theta) \end{bmatrix} \begin{bmatrix} x^* \\ y^* \end{bmatrix},$$

and $x^* \in [-w/2, w/2]$, $y^* \in [-h/2, h/2]$ is the corresponding vertex between two region tessellations.

Prop. 1 shows that for regions with identical tessellations, rotation and scaling shift correspond to the angle and scale shift of any affine identifiers in this region[3]. Building on this property, we propose URICA (Uniformity Region Affine Identifier Capture) for retrieving arbitrary regions in pathology images under varying scales and orientations. URICA uses a precomputed patch data source from $\mathcal{I}$ and a pretrained encoder $E_{s^*}(\cdot)$. Given unit patches of size s^* and tessellated queries, it estimates affine parameters, validates them, and selects the most uniform subset to retrieve candidate regions with descriptors ϕ_{res}. As shown in Fig. 3 (a), all target WSIs $I \in \mathcal{I}$ are divided into patches r_{ϕ^*} with s^* and indexed with semantic embeddings before region retrieval. Specifically, Otsu's method [35] computes the binary thumbnail I_b at 5× magnification. For each patch, given descriptor $\phi^* = (x, y, s^*, s^*, 0)$, we locate the corresponding binary patch on I_b and compute its black pixel ratio:

$$\text{ratio} = \frac{N_{\text{black}}}{N_{\text{total}}} = \frac{\sum_{x=1}^{W} \sum_{y=1}^{H} \mathbb{I}(x, y)}{H \times W},$$

where $\mathbb{I}(x, y)$ is the indicator of a black pixel: $\mathbb{I}(x, y) = 1$ if (x, y) is black and 0 otherwise. Given a threshold τ, we label a patch as foreground if its black-pixel ratio exceeds τ, and as background otherwise. Foreground patches are encoded by an $E_{s^*}(\cdot)$ and indexed to form the patch source, which supports basic patch retrieval $R_p(\cdot \mid \mathcal{I}, E_{s^*}(\cdot))$.

4.1. Uniformly Affine Identifier Computation

Fig. 3 (b) shows that when a query region r_{ϕ_q} is provided, URICA constructs a tessellation $T^t_{\phi_q} = \{V_{\phi_q}, R_{\phi_q}\}$ of patches based on the predefined step size t, and builds affine identifiers between selected patches. For a selected vertex $v_{\phi_q} \in V_{\phi_q}$ in (x^*, y^*), the top-k semantically similar patch locations are identified with the basic patch vector retrieval process:

$$\{r_{\phi^*_{res}}\}^k = R_p\left(v_{\phi_q} | \mathcal{I}, E_{s^*}(\cdot)\right),$$

where k represents the queue length of the retrieval results.

Within the tessellation, the query affine identifier p is defined by the direction from a selected vertex v_{ϕ^*} to $v_{\phi'}$,

---

[3] The complete proof is provided in Appendix C.

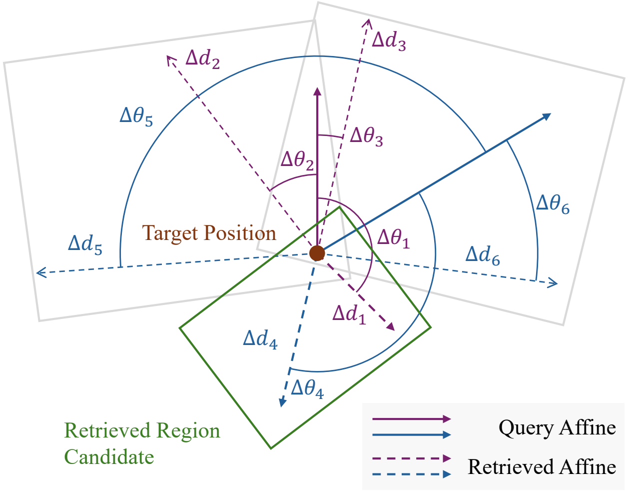

Figure 4. Correspondences between query (solid) and retrieved (dashed) affines are marked with the same color (blue and purple). The candidate (grey) and retrieved (green) regions were proposed with the shared centric position (brown) across dash combinations.

thereby encoding semantic relationships in space. The analogous identifier p_{res} is obtained from their patch retrieval results. Fig. 3 (c) shows the details.

Aggregating over all anchor pairs $v_{\phi^*} \neq v_{\phi'} \in V_{\phi_q}$, URICA estimates the shift of angle and scale ratio by:

$$p = (x' - x^*, y' - y^*),\ p_{res} = (x'_{res} - x^*_{res}, y'_{res} - y^*_{res}),$$

$$\Delta\theta = \arccos\left(\frac{p \cdot p_{res}}{\|p\| \cdot \|p_{res}\|}\right),\quad \Delta d = \frac{\|p_{res}\|}{\|p\|}, \tag{1}$$

where $(x^*, y^*), (x', y') \in \phi_q$, $(x^*_{res}, y^*_{res}), (x'_{res}, y'_{res}) \in \phi_{res}$, with $\Delta\theta$ and Δd denoting the angle and scale ratio shifts between the affine identifier.

4.2. Region Candidate Reconstruction

While Prop. 1 guarantees global consistency of transformations with rotation and scaling, URICA exploits this by validating uniformity in $\Delta\theta$ and Δd across query and retrieved regions. For each tessellation vertex pair, it computes angle and scale variation sets $\Delta\theta$ and Δd with Eq. 1, and selects consistent subsets θ^* and d^* with minimal average variance:

$$\{\theta\}^* = \underset{\{\theta\}^*, |S| \geq 2}{\operatorname{argmin}} \frac{1}{|\{\theta\}^*|} \sum_{\theta \in \{\theta\}^*} \left(\theta - \frac{\sum \theta}{|\{\theta\}^*|}\right)^2, \tag{2}$$

$$\{d\}^* = \underset{\{d\}^*, |S| \geq 2}{\operatorname{argmin}} \frac{1}{|\{d\}^*|} \sum_{d \in \{d\}^*} \left(d - \frac{\sum d}{|\{d\}^*|}\right)^2. \tag{3}$$

Then, URICA estimates the angle and scale ratio of the retrieved region relative to the query region by calculating the average angle and scale shifts over the consistent subsets:

$$\theta^* = \frac{1}{|\{\theta\}^*|} \sum_{\theta \in \{\theta\}^*} \theta,\quad d^* = \frac{1}{|\{d\}^*|} \sum_{d \in \{d\}^*} d, \tag{4}$$

where θ^* and d^* are used to locate the retrieved results.

Leveraging the shared query starting anchor (x^*, y^*) along with its validated angle θ^* and scale d^*, URICA enables the location of retrieval candidates within the center-point identifier. Fig. 4 shows that with the target point as the shared starting point, each query affine (blue and purple solid arrows) can ideally yield several retrieved affine identifiers (dashed arrows). When a subset exists such that the angle and scale shifting are uniform (arrows in green box), the angle and scale of the retrieved region candidate can be relocated. Specifically, in the region proposal step of Fig. 3 (c), the final descriptor ϕ_{res} can be computed by:

$$w_{res} = w \cdot d^*,\ h_{res} = h \cdot d^*,\ \theta_{res} = \theta^*, \tag{5}$$

$$\begin{bmatrix} x_{res} \\ y_{res} \end{bmatrix} = \begin{bmatrix} \cos\theta^* & -\sin\theta^* \\ \sin\theta^* & \cos\theta^* \end{bmatrix} \cdot \begin{bmatrix} x^* - x_c \\ y^* - y_c \end{bmatrix} + \begin{bmatrix} x^*_{res} \\ y^*_{res} \end{bmatrix}, \tag{6}$$

where (x^*_{res}, y^*_{res}) is the retrieved patch position (target point in Fig. 4) corresponding to (x^*, y^*), and (x_c, y_c) is the query region center. For a tessellation with n anchors and patch queue length k, URICA thus generates $n \times k$ candidate regions, ranks them by similarity to the query embedding, and returns the top-k most similar ones[4].

5. Theoretical Approximation Bounds

URICA uses tessellation as a sparse structure to approximate the ideal pixel-based similarity. In this section, we study their relation through a probabilistic formulation and derive bounds. First, we give a hypothesis of coincidence:

Hypothesis 1 ***(Degree of Coincidence)*** *Given an image encoder $E_{s^*}(\cdot)$ and a semantic mask $M^s_l \in \mathbb{R}^{W_l \times H_l \times d_s}$ with granularity $s > s^*$, the standard degree of semantic coincidence δ of two patches is determined by their center points, where $\forall\ x, x^* \in [0, W_l]$ and $y, y^* \in [0, H_l]$, we have:*

$$\delta = \left(\frac{2s - |x - x^*|}{2s}\right)^\alpha \left(\frac{2s - |y - y^*|}{2s}\right)^\alpha \in [0, 1],$$

where α denotes the degree of correlation.

In Hyp. 1, the parameter δ measures semantic correlation between observed (vertex) and unobserved pixels. For regions with known point-wise similarities, the bounds for unobserved points depend on semantic alignment. Given the cosine similarity $\psi \in [0, 1]$ of corresponding anchors r_ϕ at granularity s, we estimate upper and lower bounds for unobserved points using positional deviation δ_i. Aggregating across all anchors, the similarity range between two regions r_{ϕ_1} and r_{ϕ_2} with the same tessellation vertex similarities is:

$$\overline{Sim(r_{\phi'_1}, r_{\phi'_2})} = \min\left\{\overline{Sim_i(r_{\phi_1}, r_{\phi_2}|\phi)}\right\}_i,$$

$$\underline{Sim(r_{\phi'_1}, r_{\phi'_2})} = \max\left\{\underline{Sim_i(r_{\phi_1}, r_{\phi_2}|\phi)}\right\}_i,$$

---

[4] The complete algorithm is provided in Appendix D.

where each ϕ_1' and ϕ_2' denotes a local subregion centered at (a_1, b_1) and (a_2, b_2) with size $2s \times 2s$.

These bounds describe the local similarity variation inferred from the anchor-based tessellation. For the given vertices and $\forall\ (x, y) \in \mathcal{S}$, we can obtain the average interval length $L(sim(x, y))$ of the corresponding similarity as:

$$L(sim(x, y)) = E(\overline{sim}(x, y)) - E(\underline{sim}(x, y))$$

Based on the symmetry of the above probability model and a high probability, the integral $\iint_{\mathcal{S}} L(sim(x, y))ds$ can be used to quantify how closely the tessellation-based task approximates the ideal task. By guaranteeing the $\iint_{\mathcal{S}} L(sim(x, y))ds < 1 + \xi$, we can have:

$$\frac{(2s)^{4\alpha}}{(2s-t)^{4\alpha}} - \frac{(2s-t)^{2\alpha}}{(2s)^{2\alpha}} = \frac{\psi^*(1+\xi)}{2(1-\psi^*)^4} \tag{7}$$

Eq. 7 shows that, given an encoder with minimum semantic granularity s^*, the displacement similarity coefficient α can be estimated from existing WSIs. The tessellation step t is chosen so that the statistical error in comparing $\psi > \psi^*$ to segmentation-mask-based similarity does not exceed ξ[5].

6. Improving Efficiency

URICA's complexity has two main components: tessellation size $n^* = w_t \times h_t$ and retrieval queue length k. Each anchor queries the HNSW index for top-k similar patches in $O(k \log(cn))$, where c is the average number of patches per WSI, yielding a total retrieval complexity of $O(n^* k \log(cn))$. To reduce this cost, we perform anchor selection on the semantic tessellation $T_\phi^t = (V_\phi, R_\phi)$ via K-Means or spectral clustering, selecting k_a representative centers. This reduces the retrieval complexity to:

$$O(k_a \cdot k \log(cn)), \tag{8}$$

where k_a is the number of selected anchors.

Affine identification selects the subset of $k^2 n^*$ identifiers with minimal variance. Exhaustive subset enumeration yields exponential cost ($O(2^{k^2 n^*})$), so we adopt Bag of Shifting (BoS). BoS partitions angle $[-\pi, \pi]$ and scale $[0.5, 2]$ into bags (B_a, B_r), assigning each $(\Delta\theta, \Delta d)$ to intervals, with twin bags offset by half an interval to reduce boundary errors. The average variance per interval estimates (θ^*, d^*), reducing identification complexity to:

$$O(k_a \cdot k^2). \tag{9}$$

Together, anchor selection and BoS reduce the original $O(n^* k \log(cn))$ retrieval and exponential identification cost to $O(k_a k \log(cn))$ and $O(k_a \cdot k^2)$, respectively, delivering an order-of-magnitude speedup when $k_a \ll n^*$ and enabling URICA to operate efficiently on large-scale WSIs.

7. Experiments

Since URICA integrates semantic tessellation, affine identification, and theoretical bounding into a unified retrieval framework, we evaluate how each component contributes to overall performance. We therefore articulate three research questions (**RQ**) to guide our experiments:

- **RQ1**: Can URICA achieve consistent and robust retrieval performance across diverse tissue subtypes and organ systems in slide-level and region-level WSI retrieval tasks?
- **RQ2**: How do anchor selection strategies and module configurations influence the accuracy–efficiency trade-offs in URICA's multi-vertex retrieval pipeline?
- **RQ3**: How does the tessellation step, informed by the geometric bound function, influence the coverage, similarity, and retrieval latency in large-scale WSI region search?

We evaluate URICA on 24,811 TCGA[6] WSIs (13.481 TB) covering 29 cancer subtypes from 10 anatomical sites. In our experiments, WSIs are tiled into 224×224 patches at 5× magnification with step t. Patch embeddings are extracted via the UNI [7] encoder and indexed with the Milvus [30] database based on the HNSW [23] method along with spatial metadata. The experiment assesses slide retrieval for coarse matching and inter-region retrieval for fine-grained ones. Evaluation metrics include mMV@k (mean majority voting) for slides, and mSim@k (mean feature similarity) and mIoU@k (mean spatial overlap) for regions with top-k results[7].

7.1. (RQ1) WSI Retrieval Tasks Results

For experiment results, slide retrieval results are summarized in Tab. 1, and region retrieval main results in Tab. 2. Appendix F.2 gives the detailed results.

Table 1. Slide retrieval results on different subtypes of TCGA

Site	Slide Num.	mMV@5				
		Yottixel	SISH	RetCCL	HSHR	**URICA**
Pul.	3395	70.73	68.36	84.27	78.45	**99.26**
Uri.	4198	90.33	88.09	93.67	92.65	**98.80**
Mel.	1100	96.61	95.12	93.87	**97.52**	97.36
Bra.	3625	93.38	91.60	85.98	93.74	**99.81**
Gl.	3565	66.97	57.04	54.57	69.22	**97.90**
Liv./PB.	1446	89.05	86.27	89.35	91.97	**95.78**
Gyn.	3610	88.55	85.34	87.14	90.42	**98.95**
Pro./Tes.	1585	97.38	97.44	98.16	**99.10**	97.92
Hem.	421	90.38	90.51	90.71	96.40	**99.04**
End.	1866	92.49	89.80	94.31	95.19	**98.98**
Total	24811	87.73$_{\pm 09.6}$	84.96$_{\pm 12.0}$	87.00$_{\pm 13.0}$	90.87$_{\pm 10.3}$	**98.38**$_{\pm 01.2}$

To evaluate the slide retrieval task effectiveness, URICA is compared with Yottixel [14], SISH [5], RetCCL [34], and HSHR [19], using the best reported results on the TCGA dataset. Tab. 1 presents the slide retrieval performance across different cancer subtypes. The results show that URICA achieves $98.38 \pm 1.17\%$ retrieval accuracy across

---

[5] The detailed proof is provided in Appendix E

[6] https://portal.gdc.cancer.gov/

[7] More experiment details and results are provided in Appendix F

Table 2. Inter-region retrieval experiment results of different methods on different subtypes of TCGA data with top-5 Results.

WSI Site	Query	Slide Method		Sample Method		Adjacent Method		URICA (ours)		WSI Site	Query	Slide Method		Sample Method		Adjacent Method		URICA (ours)	
		mSim@5	mIoU@5	mSim@5	mIoU@5	mSim@5	mIoU@5	mSim@5	mIoU@5			mSim@5	mIoU@5	mSim@5	mIoU@5	mSim@5	mIoU@5	mSim@5	mIoU@5
Pul.	2763	0.6812	0.1656	0.8582	0.3163	0.1671	0.0262	**0.9402**	**0.6416**	**Liv./PB.**	886	0.6572	0.1505	**0.8557**	0.2788	0.1401	0.0189	0.8525	**0.5551**
LUAD	1320	0.6833	0.1673	0.8633	0.3187	0.1721	0.0287	**0.9489**	**0.6302**	CHOL	55	0.6003	0.1147	0.8493	0.2665	0.1081	0.0152	**0.9391**	**0.6251**
LUSC	1335	0.6846	0.1662	0.8536	0.3171	0.1647	0.0251	**0.9346**	**0.6576**	LIHC	520	0.6634	0.1457	**0.8653**	0.2676	0.1388	0.0178	0.8014	**0.4930**
MESO	108	0.6208	0.1432	0.8506	0.2763	0.1534	0.0177	**0.8907**	**0.5422**	PAAD	311	0.6570	0.1649	0.8409	0.2998	0.1481	0.0215	**0.8603**	**0.5821**
Uri.	3209	0.6929	0.1684	0.8602	0.3065	0.1564	0.0239	**0.9306**	**0.6485**	**Gyn.**	2562	0.6752	0.1557	0.8599	0.2960	0.1486	0.0212	**0.7993**	**0.5730**
BLCA	588	0.6325	0.1462	**0.8424**	0.2540	0.1332	0.0168	0.8402	**0.5943**	UCEC	811	0.6570	0.1424	**0.8572**	0.2556	0.1284	0.0157	0.7397	**0.5539**
KIRC	1805	0.7105	0.1748	0.8632	0.3200	0.1607	0.0254	**0.9560**	**0.6706**	CESC	358	0.5943	0.1306	**0.8443**	0.2625	0.1292	0.0154	0.6443	**0.4366**
KICH	265	0.7242	0.1815	0.8649	0.3172	0.1701	0.0287	**0.9400**	**0.5571**	UCS	95	0.6611	0.1349	0.8522	0.2780	0.1433	0.0135	**0.9601**	**0.6361**
KIRP	551	0.6846	0.1649	0.8671	0.3130	0.1607	0.0246	**0.9617**	**0.6563**	OV	1298	0.7099	0.1725	0.8664	0.3318	0.1669	0.0268	**0.9538**	**0.6764**
Mel.	645	0.6414	0.1472	**0.8422**	0.2556	0.1391	0.0168	0.7295	**0.4862**	**Pro./Tes.**	1126	0.6512	0.1633	0.8407	0.2978	0.1399	0.0221	**0.8563**	**0.5864**
UVM	97	0.6316	0.1393	0.8455	0.2227	0.1308	0.0127	**0.9584**	**0.7000**	TGCT	292	0.6614	0.1658	**0.8415**	0.3607	0.1382	0.0218	0.7296	**0.4792**
SKCM	548	0.6432	0.1486	**0.8416**	0.2614	0.1405	0.0176	0.7009	**0.4594**	PRAD	834	0.6477	0.1624	0.8404	0.2989	0.1405	0.0221	**0.9019**	**0.6250**
Bra.	2939	0.6919	0.1561	0.8727	0.2883	0.1551	0.0222	**0.9223**	**0.6427**	**Hem.**	286	0.6842	0.1605	0.8575	0.2846	0.1493	0.0194	**0.8887**	**0.6160**
GBM	1739	0.6786	0.1551	0.8649	0.2918	0.1528	0.0224	**0.8863**	**0.6117**	DLBC	77	0.6675	0.1522	0.8636	0.3112	0.1621	0.0208	**0.9550**	**0.6317**
LGG	1200	0.7112	0.1575	0.8841	0.2832	0.1585	0.0218	**0.9678**	**0.6817**	THYM	209	0.6904	0.1635	0.8552	0.2748	0.1446	0.0189	**0.8693**	**0.6114**
Gl.	2690	0.6546	0.1655	0.8456	0.3087	0.1468	0.0245	**0.8637**	**0.5953**	**End.**	1163	0.6673	0.1491	0.8564	0.2741	0.1464	0.0189	**0.8931**	**0.6393**
COAD	1140	0.6642	0.1676	0.8475	0.3132	0.1458	0.0247	**0.8573**	**0.5959**	ACC	229	0.7169	0.1647	0.8587	0.2996	0.1672	0.0197	**0.8929**	**0.6662**
ESCA	215	0.6002	0.1388	**0.8353**	0.2666	0.1202	0.0169	0.5939	**0.4923**	PCPG	242	0.6686	0.1402	0.8651	0.2619	0.1540	0.0189	**0.9387**	**0.6292**
READ	389	0.6402	0.1601	0.8483	0.3041	0.1441	0.0243	**0.9619**	**0.6160**	THCA	692	0.6505	0.1471	0.8527	0.2699	0.1368	0.0187	**0.8836**	**0.6263**
STAD	946	0.6613	0.1712	0.8444	0.3147	0.1552	0.0260	**0.8646**	**0.5959**	-	-	-	-	-	-	-	-	-	-

Table 3. Ablation experiment results in inter-region retrieval experiment of URICA on different subtypes of TCGA data.

Selection	BoS	mSim/mIoU@5										mTPQ
		Pul.	Uri.	Mel.	Bra.	Gl.	Liv./PB.	Gyn.	Pro./Tes.	Hem.	End.	
Fixed Position		0.937/0.645	**0.957**/0.653	0.756/0.454	0.927/0.626	**0.885/0.608**	0.899/0.556	**0.851**/0.570	**0.898**/0.579	0.896/0.592	0.889/0.615	24.29
	✓	0.911/0.643	0.954/**0.661**	0.678/0.447	0.902/0.613	0.845/0.594	0.897/0.568	0.825/0.570	0.866/0.572	0.876/0.589	0.871/0.620	29.60
KMeans Clustering		**0.951/0.657**	0.946/0.655	0.806/**0.519**	**0.937**/0.641	0.878/0.603	**0.918/0.609**	0.833/**0.585**	0.876/**0.590**	**0.924/0.629**	**0.934/0.646**	21.50
	✓	0.940/0.641	0.930/0.648	0.729/0.486	0.922/**0.642**	0.863/0.595	0.852/0.555	0.799/0.573	0.856/0.586	0.888/0.616	0.893/0.639	**21.41**
Spectral Clustering		0.916/0.587	0.922/0.599	**0.945**/0.441	0.867/0.544	0.849/0.553	0.817/0.478	0.770/0.491	0.847/0.540	0.832/0.505	0.851/0.543	43.15
	✓	0.830/0.529	0.868/0.560	0.581/0.363	0.817/0.517	0.787/0.510	0.709/0.434	0.660/0.426	0.795/0.496	0.749/0.485	0.780/0.498	22.99

sites, demonstrating its effectiveness in leveraging foundation models for coarse-grained tasks.

For inter-region retrieval, we compare URICA with three direct baselines: the *Slide Method*, which retrieves regions using slide-level global embeddings; the *Sample Method*, which samples candidate regions from the whole WSI and encodes them with foundation-model embeddings; and the *Adjacent Method*, which reconstructs regions from local neighboring patches. As shown in Tab. 2, URICA consistently surpasses the *Slide Method* and *Adjacent Method* by large margins on both mSim@5 and mIoU@5, and achieves the best mIoU@5 across all reported sites and subtypes. URICA also attains the best mSim@5 in most cases, with only a few exceptions where the *Sample Method* is slightly better. The gains are especially notable in pulmonary (Pul.), urinary (Uri.), and gastrointestinal (Gl.) tissues, where URICA substantially improves spatial overlap while preserving high semantic similarity. Its strong performance on smaller or more heterogeneous subtypes further suggests robust region composition under sparse or variable morphological patterns. Compared with baselines that rely on global descriptors or rigid adjacency, URICA better preserves region structure across transformations through tessellation and affine-identifier matching.

Fig. 5 visualizes region retrieval results across multiple tissue sites, showing that URICA preserves both semantic consistency and spatial morphology under transformations.

7.2. (RQ2) Ablation Study

We evaluate the impact of anchor selection strategies and module configurations through two ablation studies. First, Tab. 3 compares URICA's performance across ten tissue sites under different settings. K-Means-based anchor selection achieves the best accuracy–efficiency trade-off, outperforming Fixed Position and Spectral Clustering in inter-region retrieval, while BoS primarily benefits Spectral Clustering by reducing latency at a modest accuracy cost.

Second, Fig. 6 examines the effect of anchor ratios. Ratios below 0.5 (0.2–0.4) degrade accuracy, whereas the maximum ratio (1.0) attains peak accuracy but incurs high computational overhead. A ratio of 0.6 strikes an optimal balance, maintaining near-peak accuracy (within 4.2% loss) while substantially reducing runtime, making it well-suited for practical retrieval tasks.

7.3. (RQ3) Step of Tessellation

Based on Hyp. 1, we estimate α by sampling 100 sets of $(224, 224)$ patch pairs from 421 WSIs on hematopoietic (hem. with DLBC and THYM) across magnifications, using the UNI encoder $E_{224}(\cdot)$. For each set, we measure the cosine similarities between diagonally paired patches and derive an instance-specific α_i value based on the divergence of semantic similarity under controlled spatial offset[8].

---

[8] The estimated result is provided in Appendix F.3

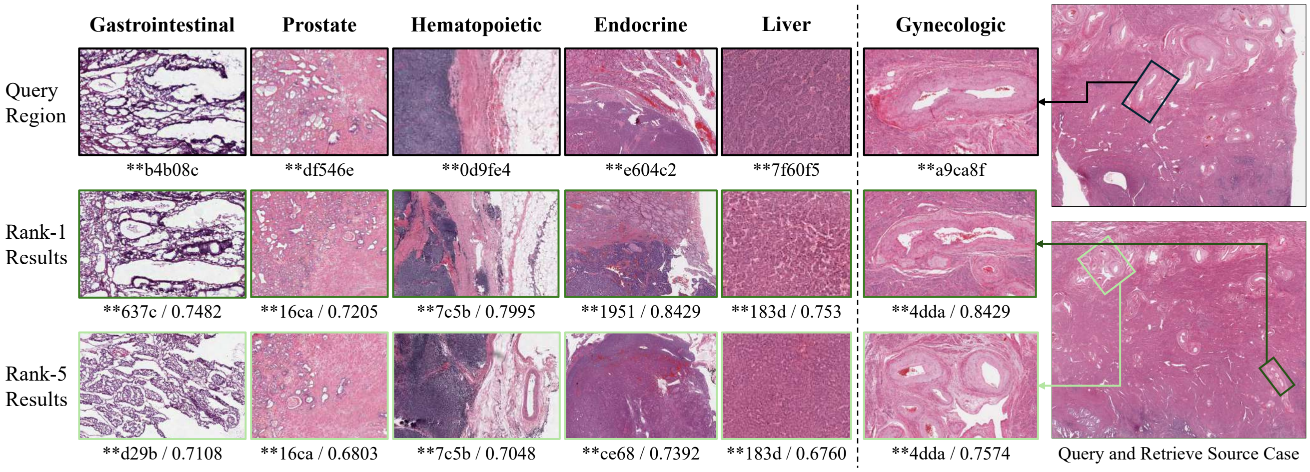

Figure 5. The rank-1 and rank-5 region retrieval results visualization of URICA across sites with original direction source illustrations.

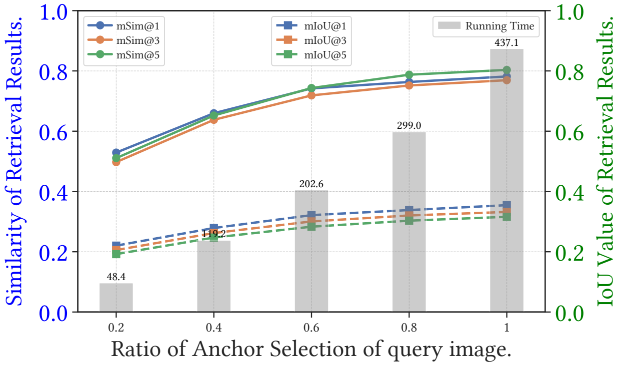

Figure 6. The differences between the region retrieval result of the full anchor and the selected anchor on TCGA.

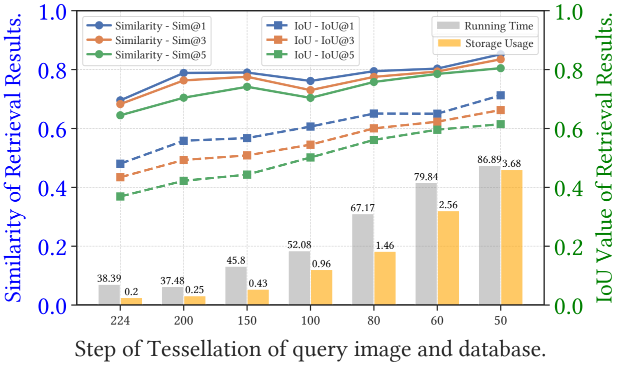

Figure 7. The region retrieval result of different tessellating steps in hematopoietic WSIs of TCGA.

The observation determined $\alpha = 3$ in Eq. 7, yielding an optimal tessellation step size of $t \approx 60$ (sampling ratio $\eta = 0.7$). As shown in Fig. 7, a large step (e.g., $t = 224$) causes sparse coverage and lower retrieval accuracy, while a small step (e.g., $t = 50$) improves accuracy but increases latency and storage. The chosen $t = 60$ achieves a good balance, reaching 0.7850 in Similarity@5 and 0.5956 in IoU@5, while being 31.6% faster and 30% more storage-efficient than the smaller-step baseline. This shows that α provides a reliable control parameter for selecting an effective tessellation granularity and validates Hyp. 1 in practice.

7.4. Limitations

While designed for region-level retrieval, URICA has three limitations. First, it is less effective for tiny or irregular regions, where tessellation and affine matching become unreliable. Second, its performance depends on encoder quality and may drop under staining variation, scanning artifacts, or rare lesion patterns. Third, it requires large-scale indexed data, and its advantage is reduced on small targets.

8. Conclusion

In this work, we presented URICA, a unified framework for WSI region retrieval that addresses the fundamental challenges of representing arbitrary regions and preserving semantic consistency under rotation and scale variations. By reformulating retrieval as semantically optimal matching, and by introducing semantic tessellation together with the affine identifier, URICA provides a transformation-robust region representation capable of composing encoder-based semantics into coherent, geometry-aware descriptors. Our theoretical analysis establishes that tessellation-based similarity tightly approximates ideal mask-level retrieval, while extensive experiments across 29 TCGA cancer subtypes and 10 anatomical sites confirm state-of-the-art performance in both slide- and region-level tasks. Beyond digital pathology, URICA offers a scalable and domain-agnostic foundation for region retrieval in other high-resolution, semantically dense imaging domains, such as remote sensing and astronomy. Future extensions will explore encoder-agnostic variants and integration into clinical workflows.

Acknowledgments

Lei Chen is supported by National Key Research and Development Program of China Grant No. 2023YFF0725100, National Science Foundation of China under Grant No. U22B2060, Guangdong-Hong Kong Technology Innovation Joint Funding Scheme Project No. 2024A0505040012, AOE Project AoE/E-603/18, Theme-based project TRS T41-603/20R, CRF Project C2004-21G, Key Areas Special Project of Guangdong Provincial Universities 2024ZDZX1006, Guangdong Province Science and Technology Plan Project 2023A0505030011, HKUST(GZ) CMCC (Guangzhou Branch) Metaverse Joint Innovation Lab under Grant No. P00659, Hong Kong ITC TC-SKLCRCC26EG01, ITF grant PRP/004/22FX, Zhujiang scholar program 2021JC02X170, HKUST Webank joint research lab. Also, this work is supported by the HKUST(GZ)-LBP Medical Data Intelligence Joint-Lab.

References

[1] Saghir Alfasly, Abubakr Shafique, Peyman Nejat, Jibran Khan, Areej Alsaafin, Ghazal Alabtah, and Hamid R Tizhoosh. Rotation-agnostic image representation learning for digital pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11683–11693, 2024. 2

[2] Shekoofeh Azizi, Laura Culp, Jan Freyberg, Basil Mustafa, Sebastien Baur, Simon Kornblith, Ting Chen, Nenad Tomasev, Jovana Mitrović, Patricia Strachan, et al. Robust and data-efficient generalization of self-supervised machine learning for diagnostic imaging. *Nature Biomedical Engineering*, 7(6):756–779, 2023. 2

[3] Georg Bökman, Johan Edstedt, Michael Felsberg, and Fredrik Kahl. Steerers: A framework for rotation equivariant keypoint descriptors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4885–4895, 2024. 2

[4] Tsai Hor Chan, Fernando Julio Cendra, Lan Ma, Guosheng Yin, and Lequan Yu. Histopathology whole slide image analysis with heterogeneous graph representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 15661–15670, 2023. 2

[5] Chengkuan Chen, Ming Y Lu, Drew FK Williamson, Tiffany Y Chen, Andrew J Schaumberg, and Faisal Mahmood. Fast and scalable search of whole-slide images via self-supervised deep learning. *Nature Biomedical Engineering*, 6(12):1420–1434, 2022. 2, 6

[6] Richard J Chen and Rahul G Krishnan. Self-supervised vision transformers learn visual concepts in histopathology. *arXiv preprint arXiv:2203.00585*, 2022. 2

[7] 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. 2, 6

[8] Linhui Dai, Hong Liu, Hao Tang, Zhiwei Wu, and Pinhao Song. Ao2-detr: Arbitrary-oriented object detection transformer. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(5):2342–2356, 2022. 2

[9] Yonghang Guan, Jun Zhang, Kuan Tian, Sen Yang, Pei Dong, Jinxi Xiang, Wei Yang, Junzhou Huang, Yuyao Zhang, and Xiao Han. Node-aligned graph convolutional network for whole-slide image representation and classification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18813–18823, 2022. 2

[10] Jiaming Han, Jian Ding, Nan Xue, and Gui-Song Xia. Redet: A rotation-equivariant detector for aerial object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2786–2795, 2021. 2

[11] Md Shakhawat Hossain, Galib Muhammad Shahriar, MM Mahbubul Syeed, Mohammad Faisal Uddin, Mahady Hasan, Shingla Shivam, and Suresh Advani. Region of interest (roi) selection using vision transformer for automatic analysis using whole slide images. *Scientific Reports*, 13(1): 11314, 2023. 1

[12] Liping Hou, Ke Lu, Xue Yang, Yuqiu Li, and Jian Xue. G-rep: Gaussian representation for arbitrary-oriented object detection. *Remote Sensing*, 15(3):757, 2023. 2

[13] Dingyi Hu, Zhiguo Jiang, Jun Shi, Fengying Xie, Kun Wu, Kunming Tang, Ming Cao, Jianguo Huai, and Yushan Zheng. Histopathology language-image representation learning for fine-grained digital pathology cross-modal retrieval. *Medical Image Analysis*, 95:103163, 2024. 2

[14] Shivam Kalra, Hamid R Tizhoosh, Charles Choi, Sultaan Shah, Phedias Diamandis, Clinton JV Campbell, and Liron Pantanowitz. Yottixel–an image search engine for large archives of histopathology whole slide images. *Medical Image Analysis*, 65:101757, 2020. 1, 2, 6

[15] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023. 2

[16] Aishik Konwer, Zhijian Yang, Erhan Bas, Cao Xiao, Prateek Prasanna, Parminder Bhatia, and Taha Kass-Hout. Enhancing sam with efficient prompting and preference optimization for semi-supervised medical image segmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20990–21000, 2025. 1

[17] Bin Li, Yin Li, and Kevin W Eliceiri. Dual-stream multiple instance learning network for whole slide image classification with self-supervised contrastive learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 14318–14328, 2021. 2

[18] Hao Li, Jiapan Wang, Balthasar Teuscher, Peng Luo, Gengchen Mai, Danfeng Hong, and Martin Werner. Gimi: A geographical generalizable image-to-image search engine with location-explicit contrastive embedding. In *Proceedings of the ICLR 2024 Machine Learning for Remote Sensing Workshop*, 2024. 2

[19] Shengrui Li, Yining Zhao, Jun Zhang, Ting Yu, Ji Zhang, and Yue Gao. High-order correlation-guided slide-level his-

tology retrieval with self-supervised hashing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9): 11008–11023, 2023. 1, 2, 6

[20] Xintong Li, Chen Li, Md Mamunur Rahaman, Hongzan Sun, Xiaoqi Li, Jian Wu, Yudong Yao, and Marcin Grzegorzek. A comprehensive review of computer-aided whole-slide image analysis: from datasets to feature extraction, segmentation, classification and detection approaches. *Artificial Intelligence Review*, 55(6):4809–4878, 2022. 1

[21] Ming Y Lu, Drew FK 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, 2021. 2

[22] Anant Madabhushi and George Lee. Image analysis and machine learning in digital pathology: Challenges and opportunities. *Medical image analysis*, 33:170–175, 2016. 1

[23] Yu A Malkov and Dmitry A Yashunin. Efficient and robust approximate nearest neighbor search using hierarchical navigable small world graphs. *IEEE transactions on pattern analysis and machine intelligence*, 42(4):824–836, 2018. 6

[24] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023. 2

[25] Andrew H Song, Richard J Chen, Tong Ding, Drew FK Williamson, Guillaume Jaume, and Faisal Mahmood. Morphological prototyping for unsupervised slide representation learning in computational pathology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11566–11578, 2024. 2

[26] Qingchen Tang, Lei Fan, Maurice Pagnucco, and Yang Song. Prototype-based image prompting for weakly supervised histopathological image segmentation. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 30271–30280, 2025. 2

[27] Wenhao Tang, Fengtao Zhou, Sheng Huang, Xiang Zhu, Yi Zhang, and Bo Liu. Feature re-embedding: Towards foundation model-level performance in computational pathology. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11343–11352, 2024. 2

[28] Hoang Vo, Yanhui Liang, Jun Kong, and Fusheng Wang. ispeed: a scalable and distributed in-memory based spatial query system for large and structurally complex 3d data. *Proceedings of the VLDB Endowment*, 11(12):2078–2081, 2018. 2

[29] Fusheng Wang, Rubao Lee, Dejun Teng, Xiaodong Zhang, and Joel Saltz. High-performance spatial data analytics: Systematic r&d for scale-out and scale-up solutions from the past to now. *Proceedings of the VLDB Endowment. International Conference on Very Large Data Bases*, 17(12):4507–4520, 2024. 2

[30] Jianguo Wang, Xiaomeng Yi, Rentong Guo, Hai Jin, Peng Xu, Shengjun Li, Xiangyu Wang, Xiangzhou Guo, Chengming Li, Xiaohai Xu, et al. Milvus: A purpose-built vector data management system. In *Proceedings of the 2021 International Conference on Management of Data*, pages 2614–2627, 2021. 6

[31] Kaibo Wang, Yin Huai, Rubao Lee, Fusheng Wang, Xiaodong Zhang, and Joel H Saltz. Accelerating pathology image data cross-comparison on cpu-gpu hybrid systems. *Proceedings of the VLDB Endowment*, 5(11):1543–1554, 2012. 2

[32] Wenyong Wang, Yuanzheng Cai, Zhiming Luo, Wei Liu, Tao Wang, and Zuoyong Li. Sa3det: detecting rotated objects via pixel-level attention and adaptive labels assignment. *Remote Sensing*, 16(13):2496, 2024. 2

[33] Xiyue Wang, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Transformer-based unsupervised contrastive learning for histopathological image classification. *Medical image analysis*, 81:102559, 2022. 2

[34] Xiyue Wang, Yuexi Du, Sen Yang, Jun Zhang, Minghui Wang, Jing Zhang, Wei Yang, Junzhou Huang, and Xiao Han. Retccl: Clustering-guided contrastive learning for whole-slide image retrieval. *Medical image analysis*, 83: 102645, 2023. 2, 6

[35] Xiangyang Xu, Shengzhou Xu, Lianghai Jin, and Enmin Song. Characteristic analysis of otsu threshold and its applications. *Pattern recognition letters*, 32(7):956–961, 2011. 4

[36] Yushan Zheng, Zhiguo Jiang, Haopeng Zhang, Fengying Xie, Yibing Ma, Huaqiang Shi, and Yu Zhao. Size-scalable content-based histopathological image retrieval from database that consists of wsis. *IEEE journal of biomedical and health informatics*, 22(4):1278–1287, 2017. 2

[37] Yushan Zheng, Zhiguo Jiang, Fengying Xie, Jun Shi, Haopeng Zhang, Jianguo Huai, Ming Cao, and Xiaomiao Yang. Diagnostic regions attention network (dra-net) for histopathology wsi recommendation and retrieval. *IEEE transactions on medical imaging*, 40(3):1090–1103, 2020. 1, 2

[38] Yushan Zheng, Zhiguo Jiang, Haopeng Zhang, Fengying Xie, Jun Shi, and Chenghai Xue. Histopathology wsi encoding based on gcns for scalable and efficient retrieval of diagnostically relevant regions. *arXiv preprint arXiv:2104.07878*, 2021. 1, 2

[39] Yushan Zheng, Zhiguo Jiang, Jun Shi, Fengying Xie, Haopeng Zhang, Wei Luo, Dingyi Hu, Shujiao Sun, Zhongmin Jiang, and Chenghai Xue. Encoding histopathology whole slide images with location-aware graphs for diagnostically relevant regions retrieval. *Medical image analysis*, 76: 102308, 2022. 2

[40] Yin Zhuang, Yuqun Liu, Tong Zhang, and He Chen. Contour modeling arbitrary-oriented ship detection from very high-resolution optical remote sensing imagery. *IEEE Geoscience and Remote Sensing Letters*, 20:1–5, 2023. 2