Bandwidth-Efficient Telepathology Viewing Framework via Detector-Guided Hybrid-Resolution Whole-Slide Image Streaming

Tong Zou[1], Tiancheng Cao[1,2], Yutong Wang[1], Han Xu[1], Hen-Wei Huang[1,3]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Department of Emergency Medicine, Brigham and Women's Hospital, Boston, USA
[3]Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore
zout0004@e.ntu.edu.sg

***Abstract*—Whole-slide images (WSIs) are gigapixel, multi-resolution pyramids that impose significant bandwidth, latency, and storage demands for interactive remote viewing. It entails a strong fidelity–efficiency trade-off: exhaustive high-resolution tile streaming preserves native details but is Input/Output (I/O) intensive, whereas aggressive compression or view-dependent resampling may improve responsiveness at the cost of visually confounding artifacts. In this work, we present a detector-guided hybrid-resolution viewing framework that transmits only a single low-resolution base level with sparse native ROIs, while reconstructing the complementary non-ROI background to yield a streamable hybrid pyramid. Our base-level tile classifier performs top-down screening to determine multi-level regions of interest (ROIs), whose indices define a determinative cross-resolution coordinate system for accurate alignment and compositing. Non-ROI background is completed by progressive residual-based super-resolution from coarse to fine levels, with hard ROI constraints to preserve diagnostic fidelity. Experimental evaluation on CAMELYON17 dataset with full-resolution WSIs as ground truth, an appropriate base level maintains a tile-level recall of 0.839 and a node-level recall of 1.000 with a reduction of 99.72% of transmitted data relative to materializing full WSI, and residual background generation reduces nuclei discrepancy by up to 54.7% compared with state-of-the-art methods. These results enable bandwidth- and storage- efficient and diagnostically faithful WSI screening for telepathology.**

***Index Terms*—Whole-slide imaging, low-resolution detection, ROI streaming, telepathology, super-resolution, cross-resolution mapping, bandwidth-efficient viewing.**

I. INTRODUCTION

Whole-slide imaging (WSI) is rapidly becoming a cornerstone of digital pathology, supporting clinical workflows such as primary diagnosis, tumor boards, and teleconsultation [1–3]. However, the multi-gigapixel, multi-resolution pyramid structure of WSIs creates a significant systems bottleneck for remote viewing. In typical clinical practice, a single compressed WSI can still reach gigabyte scale, leading to heavy bandwidth, latency, and storage demands across the client–server pipeline [4–6]. These constraints degrade interactive viewing performance and complicate long-term archival required for clinical follow-up [7, 8].

Although the DICOM-WSI standard enables interoperable tile-based access within hospital PACS systems [9], it does not resolve the core challenge of telepathology: how to provide a smooth and diagnostically faithful multi-resolution viewing experience while minimizing the transfer and storage of high-resolution data that is rarely needed everywhere at once.

This work was supported MOE RG71/24 and Nanyang Assistant Professorship. Tiancheng Cao is supported by the Wendy and Eric Schmidt AI in Science Postdoctoral Fellowship.

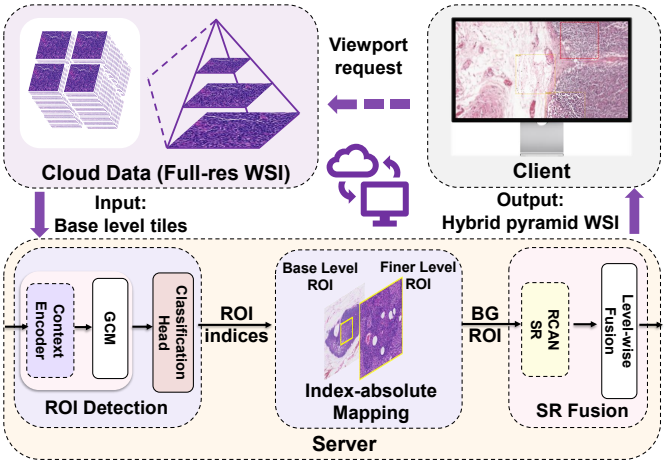

Fig. 1. **Overview of proposed framework**. Only the low resolution base level is downloaded for ROI detection. Detected ROI indices are propagated top-down to finer levels. Native ROI tiles are preserved, while non-ROI background is progressively super-resolved (×2 per level) to form a locally streamable hybrid WSI pyramid.

Existing solutions reveal a trade-off between efficiency and reliability. To reduce bandwidth and storage pressure during pan–zoom navigation, practical WSI viewers commonly rely on lossy compression and view-dependent resampling and upsampling of pyramid content, which may introduce visually misleading artifacts and alter fine-grained nuclear and texture cues for diagnosis [10, 11]. Although moderate compression can be acceptable in telepathology under certain settings, professional guidance emphasizes that compression algorithms and levels can affect the final image appearance and should be carefully validated in the intended workflow [12]. Tile-based streaming and ROI-oriented fetching reduce peak data transfer, but still rely largely on user exploration or coarse heuristics, leaving substantial background regions under-resolved and therefore provide limited context awareness: the client frequently observes isolated high-resolution (HR) "islands" over an under-resolved background, which can disrupt multi-scale reasoning and hinder reliable navigation [13, 14]. Conversely, generative reconstruction can restore visual continuity, yet unconstrained enhancement risks introducing pathology-

mimicking artifacts that undermine diagnostic trust [15, 16].

To address this, practical remote WSI reading requires a hybrid strategy that (i) detects suspicious ROIs at low resolution with minimal I/O cost, (ii) places native ROI tiles consistently across pyramid levels through deterministic cross-resolution indexing, and (iii) conservatively reconstructs only the complementary non-ROI background to maintain visual continuity[17] without introducing spurious diagnostic features.

To achieve this strategy, we present a detector-guided hybrid-resolution reconstruction pipeline for bandwidth- and storage-efficient remote WSI viewing (Fig. 1). We download only a single low-resolution base level for high-sensitivity screening and retrieve sparse native HR ROI tiles across levels via deterministic pyramid indexing, while completing the remaining field-of-view with progressive background super-resolution under ROI-preserving fusion. Overall, our contribution is a context-preserving yet bandwidth-frugal viewing scheme that maintains continuous global context, ensures ROI-consistent placement across levels, and improves visual continuity to ensure a smooth and diagnostically faithful viewing experience.

II. METHODS

A. Base-Level Detector with ROI-Driven Progressive Screening

Remote WSI viewing is fundamentally constrained by bandwidth and interactive latency, which scale with the amount of high-resolution (HR) data transmitted despite diagnostically relevant evidence being sparse. We therefore adopt a *screen-first* strategy: a lightweight detector operates on a low-resolution (LR) pyramid level to produce a conservative ROI prior, and only these ROI locations are fetched at native resolution for the subsequent cascade detection and visualization. We treat the client as a tile-request abstraction; implementation and evaluation focus on server-side construction of the hybrid-resolution pyramid.

1) Low-resolution ROI detection with Global Context Module: At the screening level, lesions may occupy only a few pixels and local texture cues become ambiguous; reliable decisions therefore benefit from global histological context. To capture such long-range context with minimal overhead, we augment a ResNet-18 backbone with a lightweight Global Context Module (GCM) as illustrated in Fig. 2. Given a LR tile input, the backbone produces a compact high-level feature map for classification. The GCM further performs global channel attention over this representation: it aggregates multi-stage feature cues from the backbone, applies channel-wise normalization to stabilize scale across stages, and then uses a shallow convolutional embedding to generate a global attention map. This attention is used to re-weight the feature map, enhancing diagnostically relevant responses while suppressing dominant background patterns. The attended features are finally fed into a lightweight classification head to predict a binary ROI logit. The detector is trained using weighted BCE-with-logits loss to mitigate the effect of imbalanced dataset.

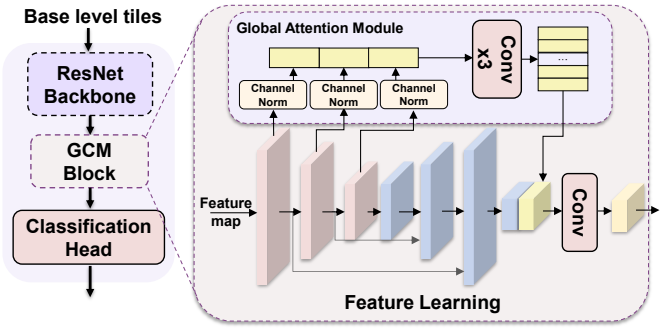

Fig. 2. **Base-level ROI detector with Global Context Module (GCM).** A 256×256 base-level tile is encoded by a ResNet backbone into a compact feature map, which is then refined by a lightweight global-attention GCM. The GCM aggregates multi-stage features with channel-wise normalization and convolutional embedding to produce content-adaptive global context, enhancing diagnostically relevant responses under low-resolution screening. The resulting representation is fed to a classification head to predict a binary ROI logit.

2) Node-balanced sampling under extreme imbalance: WSI screening exhibits severe class imbalance. To stabilize optimization, we employ a node-balanced sampling strategy. Each mini-batch first samples a fixed number of slide nodes and then draws k_{pos} positive and k_{neg} negative tiles per node whenever available. This maintains a controlled positive sampling rate across updates, improving convergence and robustness in highly imbalanced settings. At inference, the detector outputs an ROI set at the base level, which is used to index and fetch native HR tiles for the next-level cascade screening.

3) Upper-bound bandwidth and storage model: To connect screening quality to system cost, we quantify transmission and storage in level-0 tile units. A single tile at level b represents 4^b tiles at level-0. Let N_b be the number of tiles at base level b, and let $\mathrm{ROI}_b = TP_b + FP_b$ be the number of tiles marked for progressive reveal. Assuming the ROI expands by $\times 4$ per level when traversing from b to level 0 and applying no further pruning as a conservative upper bound, we define level-wise bandwidth (BW) and storage (STO) as:

$$\mathrm{BW}(b) = \frac{\mathrm{N}_b \cdot 4^b + \mathrm{ROI}_b \cdot \sum_{i=1}^{b} 4^i}{\mathrm{N}_0}, \mathrm{STO}(b) = \frac{\mathrm{ROI}_b \cdot 4^b}{\mathrm{N}_0}, \tag{1}$$

where N_0 is the total number of level-0 tiles of the WSI, N_b is the total number of level-b tiles. This formulation turns base-level selection into a trade-off: coarser screening reduces overview transmission cost ($\mathrm{N}_b \cdot 4^b$) but can inflate ROI_b through false positives, which then dominates the ROI reveal term in Eq. 1. We therefore select the base level by jointly considering detection sensitivity and the resulting bandwidth/storage upper bounds.

B. Index-Determinative Tiling for Cross-Resolution Mapping

Unlike feature-based WSI registration, we exploit the deterministic geometry of the WSI pyramid to establish an index-determinative coordinate system. A tile at level l is

uniquely mapped to a level-0 footprint via a minification factor of 2^l, so tiles across levels that cover the same tissue region share the same level-0 coordinates. This provides two practical benefits. First, ROI tiles detected at a coarse base level can be propagated to finer levels and placed back into the reconstructed pyramid deterministically without costly feature matching. Second, adjacent-level SR training pairs can be constructed by cropping HR and LR background tiles (BG–ROI pairs in Fig. 1) from the same level-0 footprint, ensuring pixel-aligned supervision for background completion.

C. Progressive Background SR with ROI-Preserving Fusion

Our pipeline reconstructs a complete yet bandwidth-efficient pyramid by synthesizing only the non-ROI background while strictly preserving native pixels inside detected ROIs. Given base-level ROI indices from Sec. II-A, we perform a top-down $\times 2$ super-resolution reconstruction from level b to level 0. For each adjacent pair $(l, l-1)$, we first synthesize the background at the finer level from the coarser-level background, and then hard-fuse native ROI tiles at level $(l-1)$ by determinative indexing. This produces a hybrid pyramid at server side that can be streamed for interactive browsing at client side under the standard tile-request abstraction, where ROIs remain native across magnifications and the complementary background is progressively refined.

1) Level-wise RCAN for ROI-aware background SR: As illustrated in Fig. 3, we adopt a level-specific Residual Channel Attention Network (RCAN) to super-resolve the background between each adjacent pyramid. Let BG_l denote the background mosaic at level l after ROI regions are masked out on the tile grid. The SR module SR_l takes BG_l as input and predicts a $\times 2$ super-resolved background at the next finer level:

$$\widehat{\mathrm{BG}}_{l-1} = \mathrm{SR}_l(\mathrm{BG}_l) = \mathrm{BG}_{l,SR}, \qquad l = b, b-1, \ldots, 1. \tag{2}$$

RCAN follows a residual-in-residual design with multiple residual groups composed of Residual Channel Attention Blocks (RCAB), and uses both long and short skip connections to stabilize optimization and preserve low-frequency structures.

2) SR training objective: For each adjacent pair $(l, l-1)$, we train a level-specific SR model using background tiles B^l. We use a distortion-oriented objective that combines an ℓ_1 term with SSIM to encourage faithful reconstruction while discouraging over-texturization:

$$\mathcal{L}_{SR} = \lambda_1 \|\widehat{\mathrm{BG}}_{l-1} - \mathrm{BG}_{l-1}\|_1 + \lambda_s \big(1 - \mathrm{SSIM}(\widehat{\mathrm{BG}}_{l-1}, \mathrm{BG}_{l-1})\big). \tag{3}$$

3) ROI-preserving super-resolution fusion: Detector-predicted ROI indices at the base level are propagated to each target level by deterministic 2×2 grid expansion, yielding an ROI mask aligned to the tile grid. At each level, we first super-resolve only the non-ROI background tiles. We then fetch the corresponding native ROI tiles using the detected index-determinative mapping and paste them back into the reconstructed canvas at the exact coordinates, overwriting any synthesized content inside ROIs. This enforces a hard ROI constraint: ROIs remain strictly native across magnifications, while SR is applied only to complementary background regions.

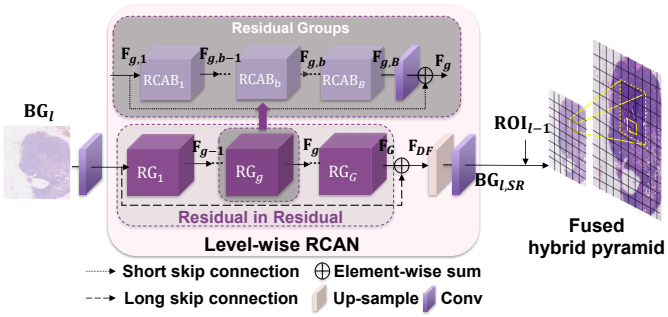

Fig. 3. **Level-wise RCAN for progressive background SR with ROI-preserving fusion.** For each adjacent pyramid pair $(l, l-1)$, a level-specific RCAN super-resolves only the non-ROI background mosaic BG_l to produce $\widehat{BG}_{l-1}$. Native ROI tiles at level $(l-1)$ are then retrieved from the full-resolution WSI repository via determinative indexing and hard-fused onto $\widehat{BG}_{l-1}$ to form a hybrid layer. Repeating this top-down $\times 2$ reconstruction yields a fused hybrid pyramid with native ROIs and progressively refined non-ROI context.

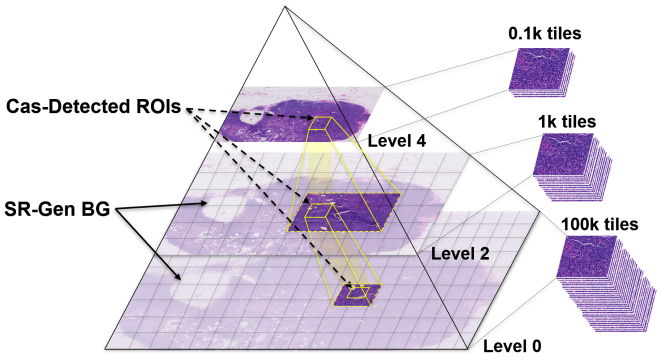

Fig. 4. **Progressive hybrid SR background and HR ROI pyramid from base level b to level 0**. Transparent regions indicate SR-generated non-ROI background, while solid regions indicate detector-identified ROIs preserved as native pixels.

4) Progressive top-down reconstruction: We generate $\hat{B}^{l-1}$ from background tiles using SR_l, fetch only native ROI tiles at level $(l-1)$, and fuse them by determinative index mapping. Repeating this iterate reconstruction for $l = b, b-1, \ldots, 1$ produces a locally browsable hybrid pyramid in which ROIs remain native across all visited magnifications, while non-ROI context is progressively refined by $\times 2$ SR. Fig. 4 depicts the progressive reconstruction from the base level b to level 0. Starting from the transmitted base layer, detected ROIs are propagated to finer levels by the deterministic grid expansion, defining where native retrieval is required. At each level, we synthesize only the non-ROI background via $\times 2$ SR, and then overwrite ROI regions with native tiles through hard fusion according to Sec. II-C3. The tile-count annotations on the right are illustrative, highlighting that while the full pyramid tile count increases quadratically toward high magnification, our transmitted native content remains sparse and ROI-driven.

III. EXPERIMENTS

A. Dataset and Training configurations

We conduct experiments on CAMELYON17 with standard train/validation/test splits and XML lesion annotations. For each slide, we construct a multi-level tile manifest using determinative level-0 footprints, enabling deterministic cross-level pairing for adjacent-level SR training and deterministic ROI mapping during reconstruction. At each pyramid level l, we extract tiles of size $P\times P$ with stride S (default $P{=}256$, $S{=}P$) and filter background using a low-resolution tissue mask; tile labels are assigned by polygon overlap with XML annotations. For hybrid reconstruction, we train a $\times 2$ SR model for each adjacent level pair $(l, l{-}1)$ using HR–LR pairs constructed from the shared index, and reconstruct progressively from the base level down to finer levels by synthesizing only non-ROI background while enforcing the hard ROI constraint to preserve native ROI pixels. Unless stated otherwise, SR is applied tile-wise with overlap-and-blend. All experiments are run on a single NVIDIA RTX 6000 PRO GPU.

For $\times 2$ background reconstruction, we compare three SR backbones with lightweight tile-wise settings (F: features; B: residual blocks; G: residual groups; A: RCABs/group; d: embedding dim; w: window): EDSR [18] ($F{=}128, B{=}24$); RCAN [19] ($F{=}32, G{=}5, A{=}5$); and SwinIR [20] ($d{=}48, D{=}(6,6,6,6), H{=}(6,6,6,6), w{=}8$). These correspond to 7.83M, 0.56M, and 0.70M trainable parameters, respectively, and we train a separate SR model for each adjacent level pair $(l, l{-}1)$.

B. Evaluation Metrics

Our metrics are aligned with the progressive reveal pipeline in Sec. II: (i) screening sensitivity, (ii) system cost in level-0 tile units, and (iii) reconstruction quality for non-ROI background SR.

1) Screening sensitivity.: We report tile-level recall and a complementary node-level recall to reflect whether lesion-containing regions are missed altogether in screening. Specifically, a positive node of a patient is counted as retrieved if any tile within that node is predicted as ROI at the screening level and thus propagated for native-resolution fetching; node-level recall is then computed over all annotated nodes.

2) System cost.: We quantify transmission and storage in level-0 tile units. A level-b tile covers 4^b level-0 tiles in field-of-view. Normalizing by the total level-0 tile count N_0, we compute conservative upper bounds using Eq. 1, where $\mathrm{BW}(b)$ assumes $\times 4$ per-level ROI expansion from b to level 0 with no further pruning, and $\mathrm{STO}(b)$ upper-bounds storing only the level-0 ROI tiles. Because the network architecture and input tile size are fixed across levels, the screening computation is dominated by the number of evaluated tiles; we therefore use $\mathrm{Cost}(b) = \mathrm{N}_b/\mathrm{N}_{L5}$ as an inference-cost proxy, where N_b denotes the number of evaluated tiles at level b and N_{L5} is the evaluated tile count at level 5.

3) Background SR quality.: For super-resolution reconstruction, we report PSNR/SSIM against native ground truth under an adjacent-level protocol, aggregated over held-out non-ROI tiles. To probe pathology-relevant structural consistency, we also report a nuclei-based cell-density [21, 22] discrepancy ΔCell:

$$\Delta\mathrm{Cell} = \frac{\left|\hat{N}(\mathrm{SR}) - \hat{N}(\mathrm{Native})\right|}{\max\left(1, \hat{N}(\mathrm{Native})\right)}. \tag{4}$$

where $\hat{N}$ is the nuclei count estimated by an off-the-shelf nuclei segmentor [23]. Because each tile has a fixed pixel size while its physical field-of-view changes systematically with magnification, $\hat{\rho}(\mathrm{Native})$ varies across pyramid levels; therefore, ΔCell values are not directly comparable across level transitions and are primarily used for within-level method comparison.

IV. RESULTS

A. Base-Level Selection via Low-Resolution Detection

Fig. 5 motivates our key claim: base-level screening efficiency is not a monotonic function of resolution. In a progressive ROI viewing, transmission is determined not only by the base-level overview, but also by how many tiles are flagged as ROI and subsequently expanded to finer levels. Therefore, the operating point must be selected from an *accuracy–bandwidth–compute* perspective rather than accuracy alone.

Fig. 5(a) shows that, as the base level becomes finer (L5→L1), tile-level recall improves because tumor evidence occupies a larger fraction of each tile and becomes more separable from background. For base-level screening, we emphasize sensitivity at the node level: at the selected operating point L4, node-level recall reaches 1.000, indicating all lesion-annotated nodes are recovered for ROI-guided reveal. At the same base level, our detector achieved tile-level recall at 0.839 under the tile-based protocol. Inference time is approximately proportional to the number of evaluated tiles, which grows sharply at finer base levels shown in the triangle curve of Fig. 5(a).

The curves in Fig. 5(b) reveal a clear optimum at L4. From L1 to L4, coarsening reduces the overview cost substantially, while lesion evidence remains sufficiently separable to keep the predicted ROI set compact; hence the bandwidth upper bound keeps decreasing. However, pushing further to L5 breaks this trend: lesions occupy a vanishing fraction of the field-of-view, separability degrades, and false positives increase. Under the conservative $\times 4$ per-level ROI growth assumption in Eq. 1, these extra FP tiles are amplified into many more level-0 tiles to fetch, producing the bandwidth rebound and dramatically increasing the false positive ROI storage in Fig. 5(b).

Table I complements the trends with concrete counts that map directly to system behavior: the confusion matrix identifies the FP surge responsible for the L5 rebound, and the storage ratio $(TP{+}FP)/N$ quantifies the practical ROI footprint that must be fetched or preserved at native resolution.

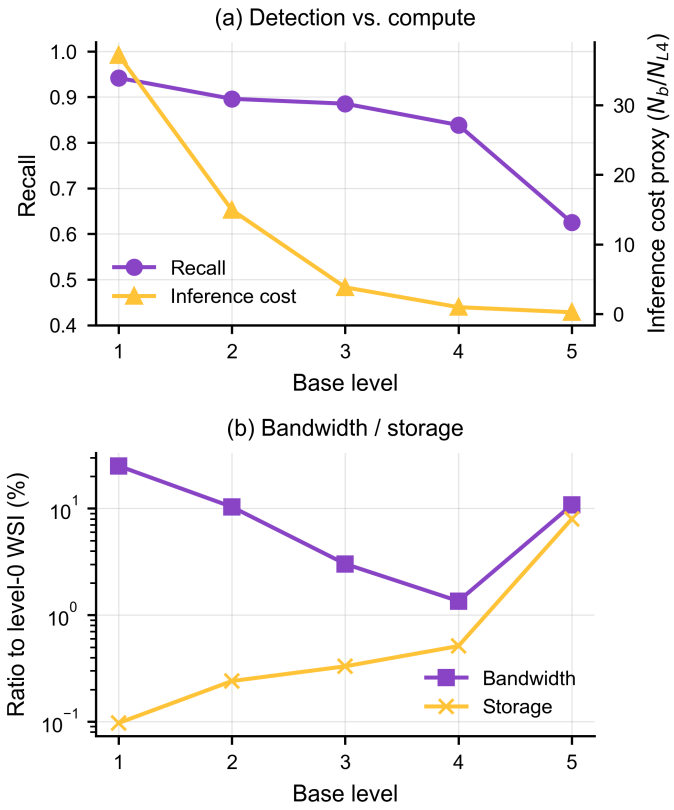

Fig. 5. **Level-wise trade-off for base-level screening.** (a) Tile-level detection recall when using level b as the screening base level and inference cost proxy shown relative to level 5 (N_b/N_{L5}) under an identical network and fixed 256×256 input tiles. (b) Upper-bound bandwidth and storage ratios, normalized to the full-resolution level-0 WSI. Bandwidth includes transmitting all base-level tiles plus the predicted ROI tiles expanded to finer levels under a ×4 per-level growth assumption and no further filtering

TABLE I
MULTI-LEVEL DETECTOR PERFORMANCE AND ESTIMATED COST PROXIES ACROSS PYRAMID LEVELS. N/N_{L5} APPROXIMATES THE RELATIVE SCREENING WITH RESPECT TO LEVEL 5. STORAGE RATIO IS COMPUTED AS $(TP + FP)/N$ (IN %).

Level	TP	TN	FP	FN	N/N_{L5}	STO ratio (%)
5	15	3655	152	9	1.00	4.3590
4	26	14405	17	5	3.77	0.2975
3	85	55486	26	11	14.52	0.1996
2	319	215716	4	37	54.60	0.1495
1	519	536742	3	3	140.24	0.0972

Together, these results support our contribution of *bandwidth-aware base-level selection* for hybrid-resolution WSI viewing. In practice, node-level recall above ∼0.9 is already within the operating range reported in clinical pathology triage studies, while extremely high sensitivity often comes at the cost of additional candidate regions to inspect [24, 25]. Our selected Level 4 offers the most stable operating point, maintaining strong sensitivity while avoiding FP-driven ROI inflation and the orders-of-magnitude compute overhead of finer screening.

B. Progressive SR for Non-ROI Background Reconstruction

We evaluate adjacent-level ×2 SR for non-ROI background reconstruction on three pyramid level transitions (L4→L3, L3→L2, and L2→L1). We compare bicubic interpolation against three representative SR backbones (EDSR [18], RCAN [19], and SwinIR [20]) trained separately for each level pair. Performance is reported on held-out non-ROI tiles using PSNR/SSIM and a risk-aware proxy ΔCell.

Table II shows that learning-based SR consistently improves PSNR/SSIM over bicubic interpolation across all level transitions and also reduces ΔCell, indicating improved fidelity without inflating nuclei statistics in typical background.

Among the compared models, RCAN yields the most consistent gains across level pairs, and we therefore adopt RCAN as our default backbone. In addition to its stable accuracy, RCAN offers a favorable system-level trade-off for interactive viewing: under our lightweight settings it uses only 0.56M parameters, which is an order of magnitude smaller than EDSR (7.83M) and comparable to SwinIR (0.70M), while avoiding transformer-specific window partitioning and padding overheads during tile-wise inference. This combination of cross-level consistency and practical deployment cost makes RCAN a robust choice for progressive non-ROI reconstruction.

Qualitatively, Fig. 6 shows that RCAN enhances local boundary definition and inter-nuclear contrast while largely preserving the spatial arrangement of nuclei compared to the native Level-1 background; the corresponding nuclei overlays exhibit similar density and layout, and in some low-contrast regions SR can make nuclei more instance-separable for the segmenter as shown in red-circled examples.

Together, these results suggest that progressive non-ROI SR can enhance contextual continuity in hybrid-resolution WSI viewing while remaining risk-aware with respect to cellular statistics, and native-resolution ROIs are preserved whenever diagnostic evidence is required.

V. CONCLUSIONS

We presented a detector-guided hybrid-resolution ROI streaming pipeline for bandwidth- and storage-efficient telepathology viewing. By downloading only a single low-resolution base level for screening and streaming native tiles only for detector-identified ROIs, our system preserves diagnostic fidelity where it matters while avoiding exhaustive high-resolution transfer. Using determinative tile indices as a cross-resolution coordinate system, we enable deterministic ROI propagation and consistent composition across pyramid levels without costly feature-based registration. For the complementary non-ROI background, we perform progressive ×2 residual super-resolution with hard ROI constraints to restore a locally browsable pyramid while suppressing visible discontinuities from synthesized content.

Experiments on CAMELYON17 support an accuracy–bandwidth–compute trade-off for base-level selection: Level 4 provides a stable operating point, achieving a patch-level recall of 0.839 with node-level recall of 1.000 and a compact native-resolution footprint with a storage ratio of 0.298%, whereas

TABLE II
NON-ROI BACKGROUND RECONSTRUCTION ON ADJACENT-LEVEL $\times 2$ SR PAIRS. METRICS ARE COMPUTED ON HELD-OUT NON-ROI TILES; n DENOTES EVALUATED TILES. HIGHER PSNR/SSIM AND LOWER ΔCELL ARE BETTER.

Pair	Method	PSNR↑	SSIM↑	ΔCell↓
L4_to_L3 (n=3725)	Bicubic	22.560	0.68847	0.15880
	EDSR	23.507	0.76887	0.11387
	RCAN	**24.327**	**0.80957**	**0.09458**
	SwinIR	23.789	0.78590	0.09539
L3_to_L2 (n=10289)	Bicubic	24.444	0.83851	0.15572
	EDSR	26.712	0.90239	0.07731
	RCAN	**27.377**	**0.91285**	**0.07060**
	SwinIR	26.590	0.89935	0.08716
L2_to_L1 (n=13765)	Bicubic	27.428	0.88498	0.11889
	EDSR	30.356	0.92809	0.09737
	RCAN	**30.735**	**0.93218**	**0.09564**
	SwinIR	30.090	0.92484	0.10627

coarser screening at Level 5 leads to FP-driven ROI inflation. For adjacent-level non-ROI background reconstruction, learning-based SR improves fidelity over bicubic interpolation by up to +3.31 dB PSNR and reduces a nuclei-based risk proxy (ΔCell) by up to 54.7%, suggesting that continuity can be improved without distorting typical cellular statistics in background tissue.

While ΔCell and nuclei-layout visual inspections provide initial evidence toward the safety of background SR, we have not yet systematically characterized the spatial distribution of false negatives or the potential for hallucinated structures in ROI-adjacent boundary regions. Future work will address these limitations by analyzing FN distributions and ROI-boundary risks via margin-based evaluations and consistency checks, while integrating realistic network and cache models for interactive viewing and conducting prospective reader studies to evaluate diagnostic confidence and workflow impact.

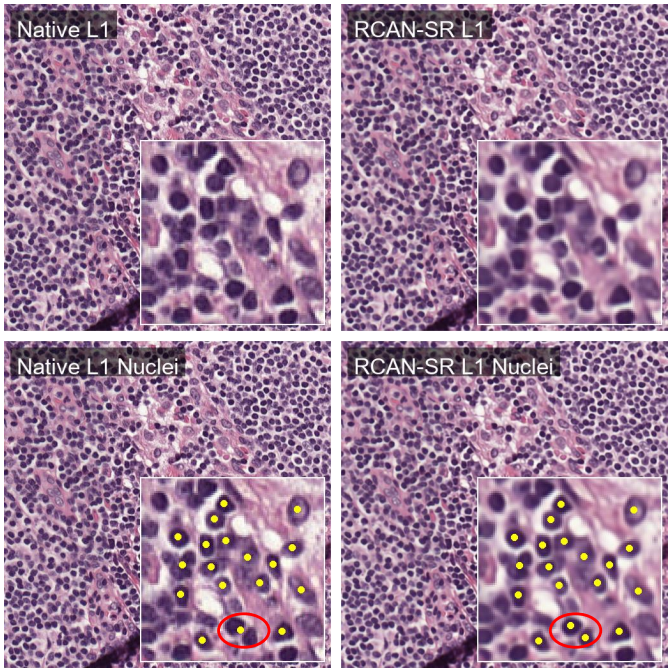

Fig. 6. **Qualitative nuclei-consistency sanity check on a non-ROI Level-1 background tile.** Top: native L1 tile and the corresponding RCAN-SR reconstruction (RCAN-SR) with an enlarged inset. Bottom: nuclei center overlays produced by an off-the-shelf nuclei instance segmentor on the inset region. RCAN-SR preserves a similar nuclei layout without introducing apparent spurious nuclei, while improving local contrast and boundary definition; as highlighted in the red circle, a low-contrast nucleus missed on the native tile becomes detectable after SR, suggesting improved instance separability for downstream nuclei analysis.

REFERENCES

[1] C. McGenity, E. L. Clarke, C. Jennings, G. Matthews, C. Cartlidge, D. D. Stocken, and D. Treanor, "Artificial intelligence in digital pathology: a systematic review and meta-analysis of diagnostic test accuracy," *NPJ digital medicine*, vol. 7, no. 1, p. 114, 2024.

[2] M. G. Hanna, V. E. Reuter, O. Ardon, D. Kim, S. J. Sirintrapun, P. J. Schüffler, K. J. Busam, J. L. Sauter, E. Brogi, L. K. Tan *et al.*, "Validation of a digital pathology system including remote review during the covid-19 pandemic," *Modern Pathology*, vol. 33, no. 11, pp. 2115–2127, 2020.

[3] S. W. Jahn, M. Plass, and F. Moinfar, "Digital pathology: advantages, limitations and emerging perspectives," *Journal of clinical medicine*, vol. 9, no. 11, p. 3697, 2020.

[4] P. Fontelo, J. Faustorilla, A. Gavino, and A. Marcelo, "Digital pathology–implementation challenges in low-resource countries," *Analytical cellular pathology*, vol. 35, no. 1, pp. 31–36, 2012.

[5] A. J. Evans, R. W. Brown, M. M. Bui, E. A. Chlipala, C. Lacchetti, D. A. Milner Jr, L. Pantanowitz, A. V. Parwani, K. Reid, M. W. Riben *et al.*, "Validating whole slide imaging systems for diagnostic purposes in pathology," *Archives of pathology & laboratory medicine*, vol. 146, no. 4, pp. 440–450, 2022.

[6] N. Kumar, R. Gupta, and S. Gupta, "Whole slide imaging (wsi) in pathology: current perspectives and future directions," *Journal of digital imaging*, vol. 33, no. 4, pp. 1034–1040, 2020.

[7] G. Velozo, J. Cordeiro, C. Nogueira, A. Teixeira, I. Costa, F. Vigatto, and F. Távora, "Implementation of digital pathology in a low-resource setting: opportunities and challenges," *Surgical and Experimental Pathology*, vol. 8, no. 1, p. 32, 2025.

[8] L. O. Schwen, T.-R. Kiehl, R. Carvalho, N. Zerbe, and A. Homeyer, "Digitization of pathology labs: a review of lessons learned," *Laboratory Investigation*, vol. 103, no. 11, p. 100244, 2023.

[9] DICOM Standards Committee, "DICOM supplement 145: Whole slide microscopic image," 2010, dICOM Standard.

[10] F. Fraggetta, V. L'imperio, D. Ameisen, R. Carvalho,

S. Leh, T.-R. Kiehl, M. Serbanescu, D. Racoceanu, V. Della Mea, A. Polonia *et al.*, "Best practice recommendations for the implementation of a digital pathology workflow in the anatomic pathology laboratory by the european society of digital and integrative pathology (esdip)," *Diagnostics*, vol. 11, no. 11, p. 2167, 2021.

[11] T. Kalinski, R. Zwönitzer, F. Grabellus, S.-Y. Sheu, S. Sel, H. Hofmann, and A. Roessner, "Lossless compression of jpeg2000 whole slide images is not required for diagnostic virtual microscopy," *American journal of clinical pathology*, vol. 136, no. 6, pp. 889–895, 2011.

[12] A. Marcelo, P. Fontelo, M. Farolan, and H. Cualing, "Effect of image compression on telepathology," *Archives of pathology & laboratory medicine*, vol. 124, no. 11, pp. 1653–1656, 2000.

[13] R. Rojansky, I. Jhun, A. M. Dussaq, S. M. Chirieleison, J. J. Nirschl, D. Born, J. Fralick, W. Hetherington, A. M. Kerr, J. Lavezo *et al.*, "Rapid deployment of whole slide imaging for primary diagnosis in surgical pathology at stanford medicine: responding to challenges of the covid-19 pandemic," *Archives of pathology & laboratory medicine*, vol. 147, no. 3, pp. 359–367, 2023.

[14] F. Ghezloo, P.-C. Wang, K. F. Kerr, T. T. Brunyé, T. Drew, O. H. Chang, L. M. Reisch, L. G. Shapiro, and J. G. Elmore, "An analysis of pathologists' viewing processes as they diagnose whole slide digital images," *Journal of Pathology Informatics*, vol. 13, p. 100104, 2022.

[15] J. Ma, J. Yu, S. Liu, L. Chen, X. Li, J. Feng, Z. Chen, S. Zeng, X. Liu, and S. Cheng, "Pathsrgan: Multi-supervised super-resolution for cytopathological images using generative adversarial network," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2920–2930, 2020.

[16] H. Lahiani, N. Khlifa, N. Navab, R. Albarqouni, and A. Kamen, "Seamless virtual whole slide image synthesis and validation using perceptual embedding consistency," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 7, pp. 2600–2610, 2021.

[17] S. Deshpande, F. A. Minhas, S. Graham, and N. M. Rajpoot, "SAFRON: Stitching across the frontier network for generating colorectal cancer histology images," *Medical Image Analysis*, vol. 77, p. 102337, 2022.

[18] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

[19] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 286–301.

[20] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 1833–1844.

[21] D. Komura and S. Ishikawa, "Machine learning methods for histopathological image analysis," *Computational and structural biotechnology journal*, vol. 16, pp. 34–42, 2018.

[22] X. Xu, S. Kapse, and P. Prasanna, "Histo-diffusion: A diffusion super-resolution method for digital pathology with comprehensive quality assessment," *arXiv preprint arXiv:2408.15218*, 2024.

[23] C. Stringer, T. Wang, M. Michaelos, and M. Pachitariu, "Cellpose: a generalist algorithm for cellular segmentation," *Nature methods*, vol. 18, no. 1, pp. 100–106, 2021.

[24] J. A. Retamero, E. Gulturk, A. Bozkurt, S. Liu, M. Gorgan, L. Moral, M. Horton, A. Parke, K. Malfroid, J. Sue *et al.*, "Artificial intelligence helps pathologists increase diagnostic accuracy and efficiency in the detection of breast cancer lymph node metastases," *The American Journal of Surgical Pathology*, vol. 48, no. 7, pp. 846–854, 2024.

[25] B. E. Bejnordi, M. Veta, P. J. Van Diest, B. Van Ginneken, N. Karssemeijer, G. Litjens, J. A. Van Der Laak, M. Hermsen, Q. F. Manson, M. Balkenhol *et al.*, "Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer," *Jama*, vol. 318, no. 22, pp. 2199–2210, 2017.