

Seamless Whole Slide Label-Free Virtual Staining

Dou Hoon Kwark¹, Kianoush Falahkheirkhah¹, Ji-Hun Oh¹, Shirui Luo²,
Volodymyr Kindratenko^{1,2*}, and Rohit Bhargava^{1*}

¹ University of Illinois Urbana-Champaign, Urbana, IL, USA

² National Center for Supercomputing Applications, Urbana, IL, USA
`dkwark2@illinois.edu`

Abstract. Label-free virtual staining offers a compelling, non-destructive alternative to standard histopathology; however, its clinical adoption is hindered by the computational bottlenecks inherent to processing gigapixel Whole Slide Images (WSIs). Current deep learning approaches require patch-based inference to avoid memory constraints, which disrupts global tissue continuity and introduces tiling artifacts—displaying visible seams and color shifts. To address this, we introduce the Consistency Memory Bank (COMB), a novel label-free virtual staining framework that enforces spatial and channel consistency across tiles without memory bottlenecks. COMB decouples context storage from computation, utilizing a dynamic retrieval mechanism to fetch feature representations from adjacent tiles. This enables a retrieval-based context integration strategy that adopts local padding to resolve spatial discontinuities and neighbor-aware channel attention to stabilize statistical drift. Further optimized with a sliding window schedule to ensure minimal memory overhead, our method demonstrates superior performance over state-of-the-art baselines, achieving significant improvements in both perceptual fidelity and tiling consistency, while suggesting its downstream utility in tumor segmentation. Code is available at <https://github.com/dou0000/COMB>.

Keywords: Label-free virtual staining · Tiling artifact mitigation

1 Introduction

Pathology diagnosis relies on physical staining to make tissue structures visible under a microscope. While effective, this standard workflow is labor-intensive, inconsistent due to chemical variations, and consumes tissue samples required for downstream molecular assays. Label-free imaging [6] offers a compelling alternative by visualizing tissue structures using their inherent signals, eliminating the need for destructive chemical dyeing process. The advent of virtual staining—using deep learning to translate raw data into histological images—offers a powerful way to optimize pathology workflows across diverse imaging modalities [22,24,7,8,9]

However, clinical adoption faces a major computational hurdle: memory constraints prevent the single holistic inference of gigapixel Whole-Slide Images

* Contributed equally

(WSIs). To that end, recent models have been developed to process the image in independent patches. This fragmentation, however, disrupts global tissue continuity, introducing *tiling artifacts*—spatial and photometric discontinuities that artificially cleave biological structures—which confound the visual continuity required for accurate pathological review.

Prior solutions for these tiling artifacts, widely adopted in virtual H&E-to-IHC staining, rely on Instance Normalization (IN) [28] augmented with global [10] or adjacent statistics [12,13]. While IN works well for standard RGB-to-RGB tasks, it can potentially underperform in label-free generation (e.g., spectral signal-to-RGB). Since it normalizes statistics based on the local statistics (in contrast to Batch Normalization (BN), which relies on learned global statistics), IN can sometimes lead to unstable predictions, particularly in label-free source modalities where signal distributions can be sparse and heterogeneous. Consequently, BN [11] remains the standard for label-free virtual staining to ensure generation quality and stability [17,27].

Despite the stability of BN, mechanisms to enforce channel and spatial consistency across patches with respect to BN remain underdeveloped. The recent CC-WSI-Net [19] attempted to address this by building upon the VSGD-Net architecture [18]—a strong generator integrating ResNet encoders and attention mechanisms—and adding global histogram matching to guide consistency and pixel-adjacency loss. However, this approach has limitations: global constraints often suppress local details [13], and simple pixel-adjacency losses are ineffective when the input domain (e.g., spectral data) lacks direct visual correlation with the output (H&E). Therefore, there is currently no single BN-based framework capable of delivering seamless, artifact-free generation for label-free virtual staining.

In this work, we introduce the **CO**nistency **ME**mory **B**ank (COMB), a novel framework designed to address the limitations of patch-based inference in label-free whole-slide virtual staining. COMB dynamically aligns adjacent patches to ensure visual continuity. Specifically, our contributions are as follows: (i) We present a novel seamless whole-slide generation framework tailored specifically for label-free virtual staining. (ii) We introduce a retrieval-based context integration strategy that integrates local padding [1] and neighbor-aware Convolutional Block Attention Module to enforce spatial and channel consistency with minimal memory overhead. (iii) We validate that our method achieves superior performance on virtual staining quality and tiling artifacts compared to state-of-the-art baselines, with demonstrated utility in downstream tumor segmentation.

2 Method

Problem Formulation. Let $x \in \mathbb{R}^{C_{\text{in}} \times H \times W}$ be a multi-spectral input and $y \in \mathbb{R}^{3 \times H \times W}$ be the corresponding H&E slide. Our goal is to learn a generator G_θ such that $\hat{y} = G_\theta(x)$ is locally faithful (preserving cellular morphology and staining) and globally seamless. To handle gigapixel WSIs, we partition x into a grid of smaller tiles $\{x_{i,j}\}_{i,j}$ of size $T \times T$ (e.g. $T = 256$ or 512).

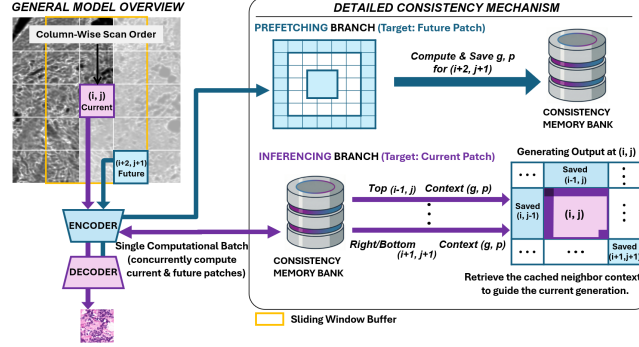


Fig. 1. COMB Architecture. A sliding-window memory bank efficiently manages global context. Described for the context radius $R = 1$

While such partition enables gigapixel WSI processing, it introduces two critical artifacts: (i) spatial discontinuities (seams) due to truncated receptive fields at tile borders, and (ii) channel-wise drift caused by the discrepancy between global training statistics and local inference windows.

2.1 Consistency Memory Bank (COMB)

To resolve these, building upon VSGD-Net [18], we introduce COMB, a novel label-free virtual staining framework that decouples context storage from computation, enabling efficient retrieval of consistent spatial and channel information to ensure seamless generation. Specifically, as shown in the model overview in Fig. 1, features from neighboring tiles are extracted by the encoder and decoder and stored in COMB. This memory bank is dynamically maintained using a sliding-window mechanism, retaining only the necessary features to avoid redundancy. During inference on a given patch, COMB retrieves the relevant neighboring features as contextual guidance, allowing the current generation to leverage consistent spatial and channel information and ensuring spatial coherence across tiles.

Spatial-Wise Consistency. Standard frameworks typically rely on explicit padding (e.g., reflection) or implicit zero-padding (e.g., the default zero-padding in standard convolution) to maintain feature map dimensions. However, these introduce synthetic boundary signals that are disconnected from adjacent patches. To address this, we implement local padding [1] via our proposed retrieval mechanism. As shown in Fig. 2(i), we explicitly pad the margins of size p by querying COMB for valid context—specifically, retrieving the boundary features from the full spatial neighborhood, utilizing stored history for processed tiles and pre-fetched representations for future context. This ensures the subsequent convolution operates on real, continuous tissue signals rather than synthetic artifacts.

Channel-Wise Consistency (Neighbor-Aware Attention). While spatial padding ensures morphological continuity, it does not correct the statistical drift

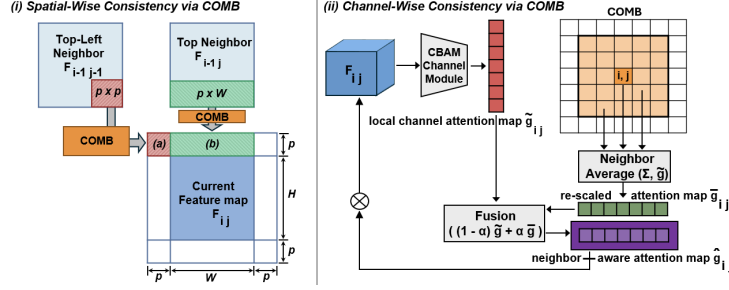


Fig. 2. Retrieved features enforce (i) spatial consistency via local padding and (ii) channel consistency via neighbor-aware CBAM.

caused by the discrepancy between the learned global training statistics and the current local inference windows. To mitigate this, we propose a neighbor-aware Convolutional Block Attention Module (CBAM) [30] integrated within the skip connections. Standard CBAM enhances virtual staining by preserving local details during encoder-to-decoder transfer [18]. However, it calculates attention weights based solely on the isolated patch, making it blind to regional inconsistencies. Consequently, it often amplifies local anomalies (e.g., stain variations) rather than correcting them. To address this, we reformulate the module to perform dynamic feature recalibration (Fig. 2(ii)). By aligning the attention mechanism with the retrieved neighbor context, we enforce regional consistency while retaining authentic tissue signals. First, we compute the standard local channel attention map $\tilde{\mathbf{g}}_{ij}$ using the isolated tile features F_{ij} via Global Average Pooling (GAP) and Global Max Pooling (GMP):

$$\tilde{\mathbf{g}}_{ij} = \sigma(\text{MLP}(\text{GAP}(F_{ij})) + \text{MLP}(\text{GMP}(F_{ij}))), \quad (1)$$

where σ denotes the sigmoid function. To harmonize the local representation with its surroundings, we compute a regional attention map $\bar{\mathbf{g}}_{ij}$ from the neighbor set $\mathcal{N}_{ij}$ (tiles within the context radius R of (i, j)) and use it to stabilize the final attention weights $\hat{\mathbf{g}}_{ij}$:

$$\bar{\mathbf{g}}_{ij} = \frac{1}{|\mathcal{N}_{ij}|} \sum_{(p,q) \in \mathcal{N}_{ij}} \tilde{\mathbf{g}}_{pq}, \quad \hat{\mathbf{g}}_{ij} = (1 - \alpha)\tilde{\mathbf{g}}_{ij} + \alpha\bar{\mathbf{g}}_{ij}, \quad (2)$$

where α is a hyperparameter. This fusion ensures that if the local signal $\tilde{\mathbf{g}}_{ij}$ is weak or anomalous, the model leverages the stable neighborhood map $\bar{\mathbf{g}}_{ij}$ to produce consistent inference.

Concurrent Prefetching and Inference Scheduling. A naive implementation of COMB requires a two-pass strategy: first caching the global context, then performing inference. However, this effectively doubles the computational cost. To overcome this, we adopt a concurrent prefetching schedule [13], enabling COMB to populate contexts in a single continuous scan. Processing in column-major order, we construct a compound mini-batch at each step containing the

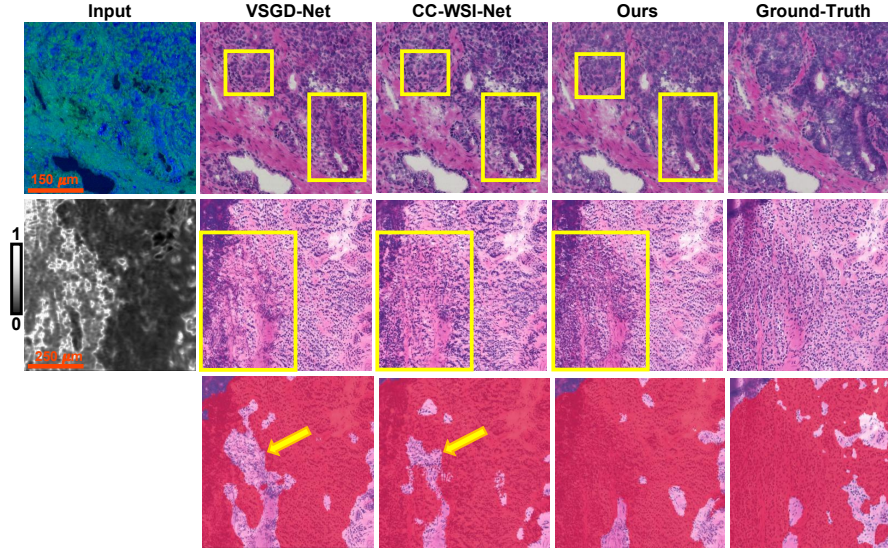


Fig. 3. Visual comparison on SRS (top) and IR (middle) datasets. The third row represents segmentation prediction on the second row.

current inference tile at (i, j) and a look-ahead prefetch tile at $(i+R+1, j+R)$, as illustrated in Fig. 1. In this setup, the prefetch tile computes and caches features into the bank, while the inference tile retrieves these pre-populated contexts to generate the final output $\hat{y}_{ij}$. This ensures the necessary neighbor contexts are available in memory just-in-time for the inference step. However, this concurrent scheduling necessitates initializing the entire global context map to accommodate the sparse updates, which is significantly expensive for gigapixel WSIs. To address this, we further optimize the memory footprint by implementing COMB as a sliding window buffer rather than a global registry. We maintain only the active columns required by the context radius R , evicting stale columns as the inference front progresses. This reduces the memory complexity from $O(\frac{H}{T} \cdot \frac{W}{T})$ to $O(R \cdot \frac{H}{T})$ in the number of cached tiles (and thus in cached feature memory), making high-resolution WSI processing feasible on standard consumer hardware.

3 Experiments

Dataset. We validated our method on a cohort of frozen prostate surgical resection samples (collected and IRB approved at Mayo Clinic), comprising two label-free modalities: multispectral mid-infrared (IR) and Stimulated Raman Scattering (SRS). The IR data was acquired using Laser Scanning Confocal Microscopy (LSCM) [4] across ten discrete wavenumbers, while the SRS [5] data consists of five spectral channels; both sets were specifically selected to capture molecular signatures of proteins, nucleic acids, lipids, and carbohydrates.

Table 1. Quantitative comparison on SRS-HE and IR-HE datasets. **Bold** indicates best, underline indicates second best. Results reported as MeanStd Dev.

Dataset	Metric	Pix2Pix	VSGD-Net	CC-WSI-Net	Ours
SRS-HE	SSIM $\uparrow$	0.4467 _{.031}	0.4809 _{.032}	<u>0.4830</u> _{.028}	0.4995 _{.043}
	LPIPS $\downarrow$	0.4413 _{.024}	0.4208 _{.012}	<u>0.4141</u> _{.016}	0.4112 _{.014}
	FSG $\downarrow$ (GT: 0.0752)	0.3960 _{.131}	<u>0.2500</u> _{.065}	0.2978 _{.096}	0.1054 _{.026}
	TexTile $\uparrow$ (GT: 0.8453)	0.3047 _{.070}	<u>0.4058</u> _{.008}	0.3743 _{.078}	0.7320 _{.066}
IR-HE	SSIM $\uparrow$	0.4846 _{.033}	<u>0.4902</u> _{.030}	0.4860 _{.026}	0.4960 _{.020}
	LPIPS $\downarrow$	0.4788 _{.004}	<u>0.4764</u> _{.008}	0.4772 _{.008}	0.4743 _{.009}
	FSG $\downarrow$ (GT: 0.1009)	0.3556 _{.036}	0.3321 _{.035}	<u>0.2669</u> _{.010}	0.1080 _{.011}
	TexTile $\uparrow$ (GT: 0.8491)	0.2500 _{.010}	0.2896 _{.010}	<u>0.3698</u> _{.003}	0.7654 _{.009}

All spectral images were spatially co-registered with their corresponding H&E stained counterparts and resized to a unified resolution of $0.5 \mu\text{m}/\text{pixel}$. For pre-processing, we discarded patches where the non-tissue area exceeded 80%. The final datasets comprise approximately 60,000 IR-HE and 16,500 SRS-HE paired 512×512 patches. For evaluation, we employed a 3-fold cross-validation strategy, ensuring strict patient-level separation between folds.

Implementation Details. We adopt the VSGD-Net [18] generator, modified to accept 10-channel IR or 5-channel SRS inputs, and a multi-scale PatchGAN [14] discriminator, optimized using the LSGAN objective [21]. For perceptual loss, we replace the standard VGG-19 [15,26] with a ConvNeXt-Tiny backbone [20] fine-tuned on histopathology data to better capture domain-specific textures. Training is performed using Adam [16] ($\beta_1 = 0.5, \beta_2 = 0.999$) with a learning rate of 2×10^{-4} and batch size of 8 for 200k iterations on an NVIDIA H200. For COMB, we set the memory bank parameters to $\alpha = 0.3$ and $R = 2$.

Experimental Setup. We benchmark our framework against three representative baselines: Pix2Pix [14] as the foundational generative model; VSGD-Net [18], a robust patch-based architecture; and CC-WSI-Net [19], a state-of-the-art virtual-staining framework for global consistency. To ensure spectral stability, we utilize Batch Normalization across all models, with the exception of CC-WSI-Net which retains its native hybrid architecture (BN encoder, IN decoder). During evaluation, WSIs are processed into 512×512 patches with a stride of 496 (16-pixel overlap). We then construct 2×2 stitched montages to center tile boundaries in the field of view. We evaluate generation quality using SSIM [29] and LPIPS [31]. In addition, we quantify stitching consistency using Focused Sobel Gradient (FSG) [2], which detects gradient discontinuities at interfaces, and TexTile [23], a learned metric assessing intrinsic seamlessness.

3.1 Results

Virtual Staining Performance. We first assess the perceptual fidelity and seamlessness of the generated H&E images on both IR and SRS modalities. As

Table 2. Segmentation performance. **Bold:** Best, Underline: Second Best (excluding GT). Results reported as Mean_{Std Dev}.

Metric	Pix2Pix	VSGD-Net	CC-WSI-Net	Ours	GT
Dice $\uparrow$	77.91 _{7.20}	79.43 _{3.35}	80.64 _{2.90}	80.97 _{4.04}	81.25 _{2.73}
IoU $\uparrow$	66.33 _{8.67}	67.33 _{4.20}	<u>69.05</u> _{3.67}	69.50 _{5.55}	70.00 _{3.54}
Sens. $\uparrow$	78.45 _{11.6}	78.13 _{2.38}	84.57 _{4.68}	<u>80.83</u> _{9.91}	84.63 _{3.52}
Spec. $\uparrow$	78.60 _{11.2}	82.08 _{8.59}	74.92 _{10.3}	<u>81.15</u> _{10.9}	75.12 _{6.54}
HD95 $\downarrow$	<u>241.37</u> _{29.0}	245.35 _{48.7}	252.25 _{47.1}	234.50 _{29.2}	224.45 _{30.9}

shown in Table 1, our framework consistently outperforms baselines across all metrics. Notably, on the seamlessness metrics (FSG and TexTile), our method shows significant improvement over other methods by a substantial margin, closely approaching the metric numbers of the Ground Truth. These quantitative results are visually confirmed in Fig. 3. In the SRS-HE translation (top-row), baselines exhibit sharp visible seams in the indicated bounding box, which leads to structural discontinuities and geometric misalignments with corresponding regions in the GT. On the other hand, in the IR-HE translation (second-row), heterogeneous signal intensities in the indicated region lead baseline generations to compromise morphological definition and obscure local tissue textures compared to our method. This validates that our retrieval-based context integration effectively resolves tiling artifacts without compromising spectral-to-morphological translation fidelity in patch-based inference.

Downstream Segmentation. To verify clinical utility, we evaluated tumor segmentation performance on the generated images (refer to rows 2-3 in Fig. 3 and Table 2), utilizing a UNet++ [3] trained on real H&E images for each fold. Consistent with our qualitative analysis, the specific artifacts identified in baseline images directly confound the segmentation model: the structural discontinuities (seams) act as artificial boundaries that fragment the prediction masks, while the uncorrected statistical drift (triggered by local signal anomalies) distorts the tumor morphology, leading to misclassification. In contrast, our seamless output restores structural integrity, enabling precise delineation of tumor regions. This is important, as it influences not only the estimation of tumor size but also the characterization of infiltrative growth patterns reflected at the tumor boundary. Inaccurate representation of these features may introduce variability in risk stratification and could potentially affect assessments of cancer aggressiveness. Quantitatively, while yielding moderate improvements in Dice and IoU, our method achieves a superior balance between Sensitivity and Specificity, avoiding the skewed performance profiles observed in competing methods. Crucially, we significantly outperform all baselines in the 95% Hausdorff Distance (HD95). As HD95 is highly sensitive to boundary outliers, this improvement confirms that our method effectively addresses the artifact-induced false positives that compromise automated analysis.

Ablation & Efficiency. We validate our architectural contributions in Fig. 4 and Table 3. As shown in Fig. 4, removing the neighbor-aware CBAM forces the

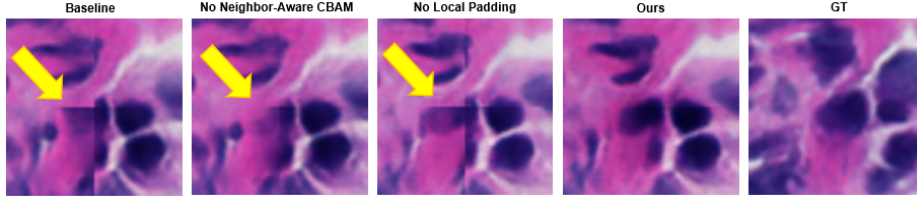


Fig. 4. Qualitative ablation study. Component-wise ablation study shows that ours effectively integrates both components for seamless, high-fidelity generation.

model to rely on individual tiles. While local padding preserves cross-patch continuity, this lack of regional context triggers channel-wise drift, leading to washed-out generations and particularly degrading the reconstruction of centrally located nuclei. Conversely, while removing local padding yields better shape reconstruction for the isolated nuclei, it immediately reintroduces boundary-localized spatial seams due to truncated receptive fields. Thus, the full COMB framework is essential for achieving both seamlessness and high-fidelity generation. Furthermore, Table 3 demonstrates the critical impact of our scheduling strategies. While concurrent scheduling significantly accelerates inference compared to the latency-heavy two-pass approach, it still shows a prohibitive memory cost. Our sliding window optimization resolves this bottleneck, reducing the memory footprint to a level close to the baseline with only a moderate increase in latency, thereby making gigapixel WSI processing feasible on standard consumer hardware.

Table 3. Inference performance on an NVIDIA RTX 2080 Ti (30720×28160 WSI).

Metric	Two-pass	Concurrent	Sliding window	Baseline
Time [s] Mem [GB]	302.4 10.64	202.6 10.64	262.8 4.38	63.9 3.50

4 Conclusion and Future Work

We presented the Consistency Memory Bank (COMB), a novel label-free virtual staining framework bridging memory constraints and global continuity in gigapixel whole-slide processing. Notably, COMB is the first fully Batch Norm-based approach to successfully resolve tiling artifacts, preserving the generation stability essential for label-free translation. By decoupling context retrieval from computation, COMB effectively mitigates these artifacts via retrieval-based local padding and neighbor-aware CBAM. Extensive evaluations on IR and SRS datasets demonstrate that our method surpasses state-of-the-art baselines in perceptual quality and generation seamlessness, and preserving this structural

continuity translates to clinical utility, significantly reducing artifact-induced errors in downstream tumor segmentation. Despite these advances, our current implementation applies local padding uniformly across the network. A promising avenue for future work involves investigating the trade-off between optimal patch dimensions and selective context retrieval. By restricting local padding to only essential architectural layers, we anticipate driving further memory optimization without compromising the seamlessness of the whole-slide generation.

Acknowledgements. This work used the Delta and DeltaAI systems at the National Center for Supercomputing Applications (NCSA) through allocation MED240056 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, and was supported in part by the Illinois Computes project (supported by the University of Illinois Urbana-Champaign and the University of Illinois System) and in part by the National Institutes of Health (NIH) under grant number R01CA260830.

Disclosure of Interests. The authors have no competing interests to declare.

References

1. Abdellatif, A., Elsheikh, A.H., Menke, H.: Local padding in patch-based GANs for seamless infinite-sized texture synthesis. arXiv preprint arXiv:2309.02340 (2023)
2. Armstrong, D., Styles, I., McLaughlin, N.: STAIN: Smooth tile-aware instance normalisation for virtual staining. In: Proc. BMVC (2025)
3. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., et al.: UNet++: A nested u-net architecture for medical image segmentation. In: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. LNCS, vol. 11045, pp. 3–11. Springer, Cham (2018)
4. Yeh, K., Sharma, I., Falahkheirkhah, K., et al.: Infrared spectroscopic laser scanning confocal microscopy for whole-slide chemical imaging. Nat. Commun. 14(1), 5215 (2023)
5. Falahkheirkhah, K., Mukherjee, S.S., Gupta, S., et al.: Accelerating cancer histopathology workflows with chemical imaging and machine learning. Cancer Res. Commun. 3(9), 1875–1887 (2023)
6. Shaked, N.T., Boppart, S.A., Wang, et al.: Label-free biomedical optical imaging. Nat. Photonics 17, 1031–1041 (2023)
7. Park, B., et al.: Rapid cancer diagnosis using deep learning-powered label-free subcellular-resolution photoacoustic histology. Sci. Adv. 11(47), eadz1820 (2025)
8. Chen, Z., Yu, W., Wong, I.H.M., Wong, T.T.W.: Deep-learning-assisted microscopy with ultraviolet surface excitation for rapid slide-free histological imaging. Biomed. Opt. Express 12(9), 5920–5938 (2021)
9. Liu, Z., et al.: Virtual formalin-fixed and paraffin-embedded staining of fresh brain tissue via stimulated Raman CycleGAN model. Sci. Adv. 10(13), eadn3426 (2024)
10. Chen, Z., et al.: Towards ultra-resolution neural style transfer via thumbnail instance normalization. In: Proc. AAAI. vol. 36, no. 1, pp. 393–400 (2022)
11. Ioffe, S., Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proc. ICML (2015)

12. Ho, M.-Y., Wu, M.-S., Wu, C.-M.: Ultra-high-resolution unpaired stain transformation via kernelized instance normalization. In: Proc. ECCV. pp. 490–505. Springer (2022)
13. Ho, M.-Y., et al.: Every pixel has its moments: Ultra-high-resolution unpaired image-to-image translation via dense normalization. In: Proc. ECCV. pp. 312–328. Springer (2024)
14. Isola, P., et al.: Image-to-image translation with conditional adversarial networks. In: Proc. CVPR. pp. 1125–1134 (2017)
15. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Proc. ECCV. pp. 694–711. Springer (2016)
16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Proc. ICLR (2015)
17. Li, J., et al.: Biopsy-free in vivo virtual histology of skin using deep learning. *Light Sci. Appl.* **10**, 233 (2021)
18. Liu, K., et al.: VSGD-Net: Virtual staining guided melanocyte detection on histopathological images. In: Proc. WACV. pp. 1918–1927 (2023)
19. Liu, S., et al.: Generating seamless virtual immunohistochemical whole slide images with content and color consistency. In: Proc. ISBI. pp. 1–5 (2025)
20. Liu, Z., et al.: A ConvNet for the 2020s. In: Proc. CVPR. pp. 11976–11986 (2022)
21. Mao, X., et al.: Least squares generative adversarial networks. In: Proc. ICCV. pp. 2794–2802 (2017)
22. Rivenson, Y., et al.: Virtual histological staining of unlabelled tissue–autofluorescence images via deep learning. *Nat. Biomed. Eng.* **3**(6), 466–477 (2019)
23. Rodriguez-Pardo, C., et al.: TexTile: A differentiable metric for texture tileability. In: Proc. CVPR. pp. 4439–4449 (2024)
24. Schnell, M., et al.: All-digital histopathology by infrared–optical hybrid microscopy. *PNAS* **117**(7), 3388–3396 (2020)
25. Bhargava, R., et al.: Digital histopathology by infrared spectroscopic imaging. *Annu. Rev. Anal. Chem.* **16**, 205–230 (2023)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: Proc. ICLR (2015)
27. Soker, A., et al.: AI-powered spectral imaging for virtual pathology staining. *Bio-engineering* **12**(6), 655 (2025)
28. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
29. Wang, Z., et al.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)
30. Woo, S., et al.: CBAM: Convolutional block attention module. In: Proc. ECCV. pp. 3–19 (2018)
31. Zhang, R., et al.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proc. CVPR. pp. 586–595 (2018)