Parameter-Efficient Fine-Tuning of Pathology Foundation Models for Microsegmentation of Breast Whole-Slide Images in Digital Pathology

M.A. Al Mamun, S. Purba, I. Obeid and J. Picone

The Neural Engineering Data Consortium, Temple University, Philadelphia, Pennsylvania, USA
{abdullah18, sadia.afrin.purba, iobeid, picone}@temple.edu

***Abstract*— Microsegmentation of high-resolution whole-slide digital pathology images remains a challenging problem, limited less by model capacity than by the scarcity of densely annotated training data. Using the Temple University Hospital Digital Pathology Breast Tissue Corpus, we ask whether pathology foundation models, adapted with low-rank adaptation (LoRA), improve performance compared to a convolutional neural network trained from scratch. Three foundation models – GigaPath, UNI, and CONCH – are evaluated in two settings: frozen linear probing and LoRA fine-tuning and compared against an EfficientNet-B0 baseline. Each frozen foundation model exceeds the performance of the baseline, with additional gains provided by LoRA. The best overall system, CONCH combined with LoRA, achieved an MDICE score of 0.646 and a Cohen's κ of 0.544, representing an 11.20% relative improvement over the baseline. We also introduce polyform, a label-preserving region former that turns the frame-prediction grid into valid, non-overlapping polygons, substantially outperforming the conventional relabeling post-processors used in the baseline system.**

***Keywords*— microsegmentation, foundation models, low-rank adaptation, digital pathology, whole-slide imaging**

1. Introduction

Digital pathology is transforming tissue analysis [1], but it also presents new challenges for pathologists, who must review vast amounts of high-resolution image data. A single hematoxylin-and-eosin (H&E) whole-slide image (WSI) can exceed ten billion pixels, and a breast case may span several such slides. Microsegmentation – the assignment of a tissue class to every region of a slide, rather than a single label to the whole image – is the form of automation most useful to a reporting pathologist, because it localizes invasive carcinoma, ductal carcinoma in situ (DCIS) and their benign mimics [2]-[4]. It is also the most challenging task for model development because it requires detailed manual annotations. Though applications of deep learning to digital pathology have advanced from metastasis detection [5] to weakly-supervised whole-slide classification [6][7] region-level microsegmentation using partial annotations remains a challenge.

Two properties of the problem shape everything that follows. First, annotation is partial: exhaustively outlining every region of a gigapixel slide is impractical, so expert labels cover only a fraction of the tissue. On a typical slide in the TUH Digital Pathology Breast Tissue Corpus (TUBR) [3], less than 2% of the tissue area is annotated. A model is therefore trained and scored on islands of ground truth in a sea of unlabeled tissue, which distorts any metric that counts predictions falling outside those islands. Second, the appearance of breast tissue is enormously heterogeneous, and a convolutional network trained from scratch on a limited pool of labelled frames tends to overfit the textures it has seen rather than learn morphology that transfers across patients and scanners.

Foundation models (FMs) offer a way around the second problem. Following the vision-transformer [8] and self-distillation [9] recipes that proved effective on natural images, encoders pretrained by self-supervision on tens of millions to billions of unlabeled histology tiles (UNI [10], CONCH [11], Prov-GigaPath [12], Virchow [13] and CTransPath [14]) learn representations that transfer broadly across organs and tasks. Adapting such an encoder to a new task by full fine-tuning is expensive and prone to overfitting. Low-rank adaptation (LoRA) [15] instead freezes the pretrained weights and learns a small pair of low-rank matrices per adapted layer, changing well under 1% of the parameters. It has become the default parameter-efficient fine-tuning (PEFT) method in language modeling [16]. It has also recently been applied to adapt a multimodal language model for patch-level classification on this same breast corpus [17]. LoRA's behavior on vision-only pathology encoders for dense microsegmentation remains uncharacterized.

Earlier work established an EfficientNet-B0 (EB0) [18] microsegmentation system as state of the art on this task. Approaches such as residual networks [19], U-Net [20] and HoverNet [21] must learn histological morphology from a limited pool of labeled tissue. Optimum performance is in the mid-50 percentile for the MDICE metric [2]. The question we pose here is whether a LoRA-adapted pathology FM can achieve superior performance when all three are decoded and scored through an identical pipeline?

There are two main contributions in this paper: (1) we report a controlled comparison of three pathology FMs, demonstrating the efficacy of the LoRA approach; and (2) we introduce polyform, a label-preserving region former that converts a frame-prediction grid into valid, non-overlapping polygons without the accuracy loss of the conventional relabeling post-processors. We also are providing the entire system, including the training protocols, as an open source toolkit so that our results can be easily reproduced.

2. Data

All experiments use TUBR v5.0.0 [3]. Annotations use nine labels: background, null, artifact, non-neoplastic, inflammation, suspicious, invasive carcinoma, ductal carcinoma in situ, and normal. The diagnostically salient tissue classes together with background are scored. Annotations of these events are partial by design – only a fraction of the slide area has been manually labeled. Because a large portion of the slide is unlabeled, any false-alarm rate we report is an upper bound rather than a measurement (see Section 3.6).

The corpus is divided by slide into 1,652 training, 928 development and 921 evaluation slides. Three slides from the development test set (/dev) that were completely annotated were held out of the per-epoch selection to avoid skewing the F1 score. These slides were reserved as a densely-labeled control for direct estimation of the false alarm rate as Section 3.6. The data split is the same for every system, so any difference in score reflects the model rather than the data it saw.

Table 1 summarizes the corpus. Table 2 gives the per-class region counts over all three splits. The corpus is strongly unbalanced, and the diagnostically critical carcinoma classes are the rarest – which motivates the class-balanced sampling described in Section 4. TUBR v5.0.0 is also a substantially refined re-annotation of the corpus. On the shared evaluation set the region count rises from 6,259 in v4.0.0 to 14,117 in v5.0.0, a 2.26x increase in annotated regions. Hence, v5.0.0 is a fairer and more demanding basis for comparison.

Table 1. An overview of TUBR

Split	Slides	Role
Train	1,652	encoder / head adaptation
Dev	928 (+3 dense)	model selection (dev macro-F1)
Eval	921	held-out reporting

Table 2. Per-class annotated region counts

Class	Regions	%
Non-neoplastic	14,269	30.60
Normal	13,312	28.60
Background	6,785	14.60
Null	2,904	6.20
Inflammation	2,700	5.80
Invasive carcinoma	2,552	5.50
Artifact	2,176	4.70
DCIS	1,673	3.60
Suspicious	237	0.50

3. Methods

In this section, we describe the processing pipeline, the fine-tuning approaches and the evaluation methodology.

3.1. Frame-Level Decoding

Each slide is tiled into non-overlapping 256 x 256 pixel frames. Tissue-bearing frames are selected with an Otsu threshold on a downsampled thumbnail followed by morphological cleaning [22], so that glass background is never decoded. Every frame is passed to the encoder together with a surrounding context window of $512\,x\,512$ pixels, and the encoder emits a nine-way posterior for the frame. The result for a slide is a grid of per-frame class posteriors. This front end is identical for all systems, so that differences in the final score are attributable to the encoder and its adaptation rather than to tiling or tissue detection. This is shown in Figure 1.

3.2. Foundation Model Encoders

We evaluate three pathology foundation models that span the current design space. UNI [10] is a ViT-Large encoder trained by self-distillation on over 100K whole-slide images and produces a 1,024-dimensional embedding. CONCH [11] is a vision language model whose 768-dimensional vision tower was trained on curated image-caption pairs. Prov-GigaPath [12] is a tile encoder trained on more than one billion tiles that yields a 1,536-dimensional embedding. In each case we use only the tile-level encoder and discard any slide-level aggregator, since our unit of prediction is the frame.

3.3. Frozen Probing and LoRA Adaptation

We compare two ways of using each encoder. In the frozen regime the encoder weights are fixed and only a small classification head is trained on the frozen features, a linear probe with one hidden layer. In the LoRA regime the encoder is adapted end-to-end with the head. For a

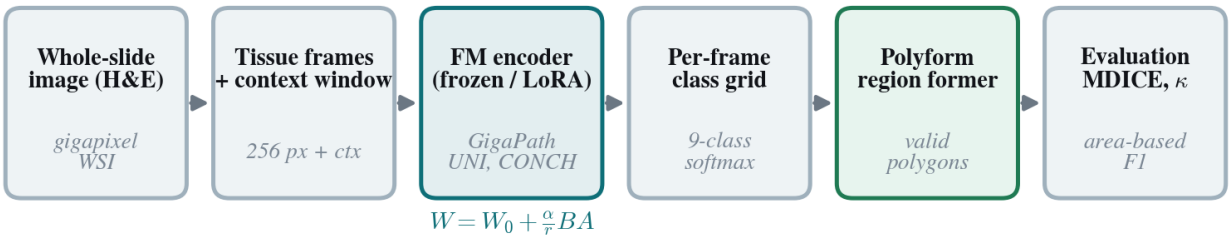

Figure 1. The microsegmentation pipeline, from whole-slide tiling through a frozen or LoRA-adapted foundation-model encoder to polyform polygons scored by MDICE and κ.

pretrained projection weight, W_0, LoRA leaves W_0 untouched and adds a low-rank update, so that the adapted layer computes

$$h = W_0 x + \frac{\alpha}{r} BAx, B \in \mathbb{R}^{d\times r}, A \in \mathbb{R}^{r\times k}, r \ll min(d,k) \quad (1)$$

where only A and B are trained and the rank r is far smaller than the layer dimensions; we use $r = 8$ and $\alpha = 16$, injected into the attention projections. The frozen probe is the $r = 0$ special case in which the encoder does not move at all. Both regimes optimize a class-weighted cross-entropy over the labelled frames,

$$\mathcal{L} = -\frac{1}{|\mathcal{M}|}\sum_{i\in\mathcal{M}} w_{y_i} log\, \hat{p}_{i,y_i}, \quad (2)$$

where M is the set of frames whose reference label is known, the weights counter class imbalance, and the best checkpoint is chosen by macro-F1 on the development set. Because LoRA leaves the pretrained weights in place and learns only a low-rank correction, it is expected to preserve the encoder's transferable features while nudging them toward the task. Whether it does so in practice is one of the questions we evaluate in this work.

3.4. A Convolutional Neural Network Baseline

Our baseline is an EfficientNet-B0 [18] trained from scratch on the same frames using the identical decoding, training, and evaluation pipeline. It represents the conventional convolutional neural network (CNN) approach and reproduces the benchmark system previously established for this corpus [2]. The baseline is decoded and evaluated identically to the foundation models to ensure a fair comparison. Because LoRA is not applicable to CNNs, all EfficientNet-B0 weights are optimized during training.

3.5. Polyform Region Forming

The output from the first pass of decoding is a grid of frame-level hypotheses on the order of several thousand per slide. Two problems make it unfit as a deliverable. First, the small squares do not correspond to the boundaries of the regions of interest. Second, adjacent squares of the same class are not merged. A naive union can produce overlapping geometry that inflates any area-based score. Conventional post-processors [2] such as flood-fill by class priority, k-nearest-neighbor voting, and modal smoothing, were designed to clean noisy output. Because the quality of our FM's predictions has improved, we are able to employ a different postprocessing strategy.

The polyform algorithm instead forms regions without ever changing a label. It takes the 8 adjacent connected components of frames that share a class, computes the exact union of each component's frame boxes into a single polygon, and discards components smaller than a minimum size. This is illustrated in Figure 2 for a carcinoma region that encloses a small DCIS region.

Because our annotation format stores one label per region and cannot represent a hole, a region that encloses an island of another class is split by a straight cut through the enclosed region rather than having the hole filled. Filling would relabel the enclosed region into the surrounding class. Each frame belongs to exactly one label, so the regions are disjoint by construction, and the output is a set of valid, non-overlapping, hole-free polygons.

3.6. Evaluation

Scoring of decoder output follows a rigorous process defined in Hackel et al. [2][23]. For each reference region, the fraction of its area covered by each hypothesis region is computed and aggregated into a class confusion matrix:

$$M[c_r, c_h] \leftarrow M[c_r, c_h] + \frac{|R_r \cap R_h|}{|R_r|}. \quad (3)$$

Precision, recall and F1 scores follow. Here R_r and R_h are a reference and a hypothesis region and c_r, c_h their class labels. Because the hypothesis regions are non-overlapping, no area is double-counted. This metric is referred to as a modified DICE score (MDICE). It is the macro-average of the per-class F1 scores:

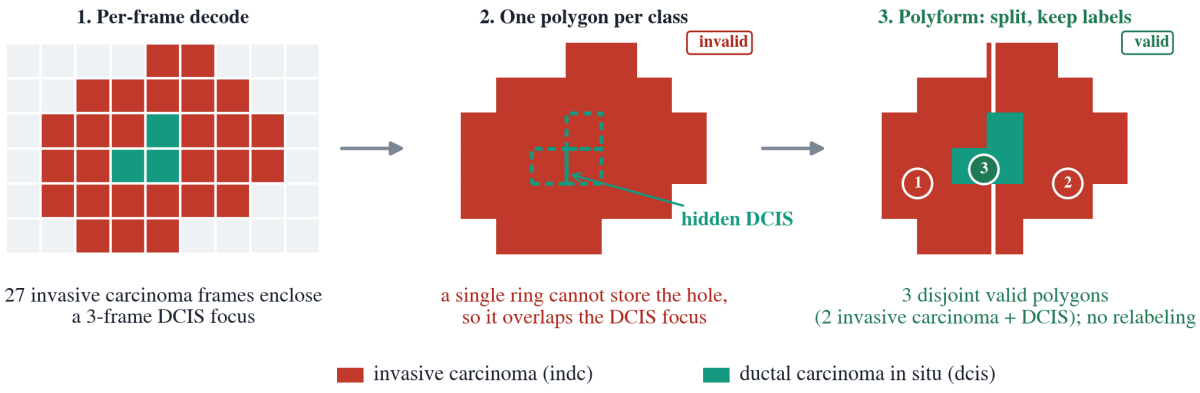

Figure 2. A demonstration of how the polyform algorithm splits an invasive carcinoma region that encloses a DCIS focus into valid, non-overlapping polygons without relabeling any frame (right).

$$F1_c = \frac{2\ TP_c}{2\ TP_c + FP_c + FN_c}, \quad (4)$$

$$MDICE = \frac{1}{|C|}\sum_{c \in C}\ F1_c\ . \quad (5)$$

We report Cohen's κ [24] alongside MDICE as a chance-corrected measure of agreement,

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \quad (6)$$

with p_o the observed and p_e the expected agreement. Both metrics are computed only where reference labels exist. A prediction on unlabeled tissue is neither rewarded nor penalized, so agreement metrics of this kind are fair for partially annotated data, and any false-alarm count is an upper bound rather than a measurement.

4. Experimental Design

All models are trained with AdamW [25] using a weight decay of 10^{-4} and a batch drawn by class-balanced sampling capped at 8,000 frames per class, with a frame counted as belonging to a class when at least 80% of its area is covered by that region. The shared LoRA recipe uses a learning rate of 10^{-4} for six epochs. The best checkpoint is selected by the development macro-F1 score. The frozen probe trains only the head under the same schedule. Decoding is identical across systems, and post-processing uses polyform with a two-frame minimum region size unless a legacy comparison is stated. Table 3 lists the settings.

Whole-pipeline decoding of a median slide (~8,500 tissue frames) takes between $91s$ and $207s$ on a single A40 GPU, as shown in Table 4. The computational cost is dominated by slide I/O and preprocessing rather than the encoder: CONCH-LoRA, UNI-LoRA and GigaPath-LoRA are about $2.3x$, $1.6x$ and $1.8x$ slower than EB0. Decoding is a one-time, offline step.

Table 3. Our common system configuration

Component	Value
Frame / context	256 px frame, 512 px content window
Encoders	GigaPath, UNI, CONCH; EB0 baseline
LoRA rank / alpha	8 / 16, attention projections
Optimizer	AdamW, weight decay 10^{-4}
Learning rate / epochs	10^{-4} / 6 (dev-F1 selection)
Sampling	class-balanced, ≤ 8,000 / class
Label threshold	≥ 80% frame coverage
Post-processing	polyform, min. 2 frames

Table 4. Whole-pipeline decode time per slide (a single A40 GPU; mean time computed over 10 medium-sized slides)

System	Encoder	Context window	Decode (s/slide)
EB0	EfficientNet-B0	512 px	91
UNI-LoRA	UNI ViT-L	512 px	147
GigaPath-LoRA	Prov-GigaPath ViT	512 px	160
CONCH-LoRA	CONCH vision tower	512 px	207

5. Results and Analysis

In this section, we present a comparison of FMs to the baseline model, discuss some relevant issues with fine-tuning, and evaluate postprocessing strategies.

5.1. Foundation models versus the baseline

In Figure 3 and Table 6, we report MDICE and κ for every encoder in both regimes, together with the convolutional baseline, on the 921-slide evaluation set. The first observation is that the foundation models are strong even without adaptation: all three frozen probes, between 0.600 and 0.616 MDICE, exceed the EB0 baseline at 0.581. A linear read-out of pretrained pathology features already outperforms a convolutional network trained end-to-end on the same labeled frames, which is direct evidence that the pretrained representations transfer to microsegmentation.

The second observation is that adaptation helps, and the best system is CONCH with LoRA at 0.646 MDICE and 0.544 κ, an 11.2% relative improvement over the baseline. GigaPath with LoRA follows at 0.635. The ranking is consistent between MDICE and the chance-corrected κ, which guards against the possibility that the macro-F1 gain is an artifact of class prevalence.

The gains are statistically significant, as shown in Table 7. A slide-level bootstrap (5,000 resamples of the 921 evaluation slides, re-aggregating the area-based confusion matrix each time) gives a corpus-MDICE improvement over EB0 of $+0.068$ (95% CI $[0.050, 0.086]$) for CONCH-LoRA, $+0.057$ $[0.040, 0.075]$ for GigaPath-LoRA, and $+0.026$ $[0.009, 0.046]$ for UNI-LoRA (all $p \leq 0.003$). None of the intervals include zero.

An analysis of the FM advantage is presented in Table 8. The performance gains are concentrated in invasive carcinoma (indc: 0.79-0.86 vs 0.63) and inflammation

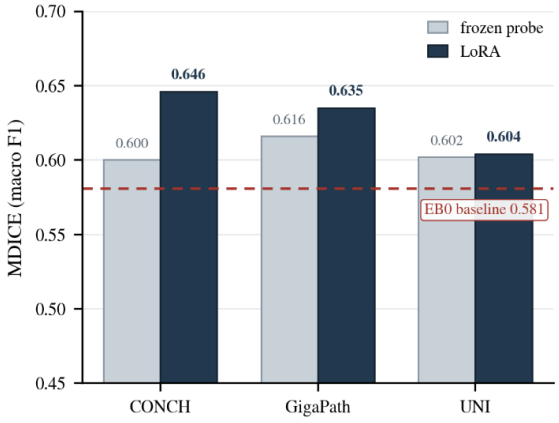

Figure 3. MDICE scores by encoder in the frozen and LoRA regimes compared to the EB0 baseline (dashed).

Table 6. Frozen versus LoRA MDICE and Cohen's κ on the 921-slide evaluation set, after polyform.

Encoder	Frozen	κ	LoRA	κ
CONCH	0.600	0.487	**0.646**	0.544
GigaPath	0.616	0.502	**0.635**	0.529
UNI	0.602	0.486	**0.604**	0.486
EB0 (CNN)	—	—	**0.581**	0.477

Table 7. A statistical significance analysis of the FM gains

System	Δ MDICE vs EB0	95% CI	P
CONCH-LoRA	+0.068	[0.050, 0.086]	< 0.001
GigaPath-LoRA	+0.057	[0.040, 0.075]	< 0.001
UNI-LoRA	+0.026	[0.009, 0.046]	< 0.003

Table 8. Per-class F1 with 95% bootstrap confidence intervals (921-slide evaluation, after polyform)

Class	CONCH-LoRA	GigaPath-LoRA	UNI-LoRA	EB0
nneo	0.640 [.62,.66]	0.574 [.56,.59]	0.365 [.34,.39]	0.644 [.63,.66]
infl	0.567 [.53,.61]	0.612 [.57,.65]	0.564 [.52,.61]	0.441 [.39,.49]
dcis	0.584 [.51,.65]	0.533 [.46,.60]	0.614 [.54,.68]	0.513 [.44,.57]
indc	0.793 [.74,.83]	0.807 [.76,.85]	0.861 [.82,.89]	0.633 [.56,.69]
bckg	0.657 [.64,.67]	0.664 [.65,.68]	0.631 [.62,.65]	0.673 [.66,.69]

(infl: 0.56-0.61 *vs* 0.44); on non-neoplastic (nneo) tissue and background (bckg), the CNN baseline is comparable.

5.2. Adaptation is encoder-dependent

The single shared LoRA recipe that improved CONCH and GigaPath does not transfer unchanged to every backbone. Applied to UNI, already the strongest frozen encoder at 0.602, the shared recipe drove it below its own frozen probe. Adapting more gently per encoder (learning rate 10^{-5}, rank 4) recovers that loss and reaches 0.604, matching the frozen probe rather than improving on it. UNI's pretrained features are evidently already well aligned to the task, leaving little room for adaptation to help. The practical guidance is to match the adaptation strength to each encoder rather than assume one shared schedule transfer.

5.3. Region forming and post-processing

In Table 5, we isolate the effect of the region former, holding the encoder fixed. The legacy relabeling post-processors, priority and k-nearest-neighbor, sit in the 0.40-0.48 range for every model, because they rewrite roughly one frame in seven. This erodes the clean predictions the encoder produced. Polyform, which never changes a label, scores far higher: 33% above the legacy methods, averaged over the models, while producing the valid, non-overlapping polygons a pathologist can read. The clinical rationale for the polygonal representation builds on our earlier work on automated interpretation and visualization in digital pathology [1].

Table 5. Post-processing algorithm comparisons

Model	Priority	kNN	Polyform
CONCH-LoRA	0.483	0.483	**0.646**
GigaPath-LoRA	0.479	0.476	**0.635**
EB0	0.437	0.436	**0.581**

It is worth stating why filling an enclosed island, rather than splitting, would be the wrong choice. Merging a small, enclosed region into its surrounding class to make a hole-free polygon relabels those frames. When the island is a correct minority prediction, such as DCIS inside invasive carcinoma, relabeling erases a correct call, Because MDICE is a macro-average over classes, the minority loss is not diluted. Splitting the region instead keeps every label and remains valid, which is why polyform preserves the encoder's accuracy rather than trading it for a smoother shape.

6. Discussion

The evidence supports a measured version of the foundation model thesis. Pretrained pathology features are strong enough that even a frozen encoder with a linear probe beats a CNN. LoRA adaptation provided further gains, especially when coupled with CONCH. The size of that gain is encoder dependent: a single shared recipe fits some backbones better than others. For UNI, a gently tuned per-encoder LoRA only matches its frozen probe, since UNI's features are already well aligned to the task. The practical guidance is to match the adaptation strength to each encoder rather than assume one schedule transfers, and the frozen results already give a strong, uniform baseline to build on.

The post-processing study makes an important second point: how the frame grid is turned into regions matters as much as which encoder produced it. A relabeling step designed for noisy convolutional output silently erases the advantage of clean FM predictions, while a label-preserving polyformer retains it and yields clinically valid polygons at no cost in accuracy. For a system whose output a pathologist is meant to read, that distinction is not cosmetic. This is demonstrated in Figure 4.

Several limitations qualify these findings. Partial annotation bounds the false alarm analysis. Because unlabeled predictions are not scored, detection-style error rates are only upper bounds. A densely annotated control set is needed to measure them directly. The results are from a single institution's breast corpus, and the UNI adaptation is not yet tuned per encoder. None of these caveats changes the central ordering: adapted foundation models lead, frozen foundation models beat the convolutional baseline, and label-preserving region forming is what makes the predictions usable.

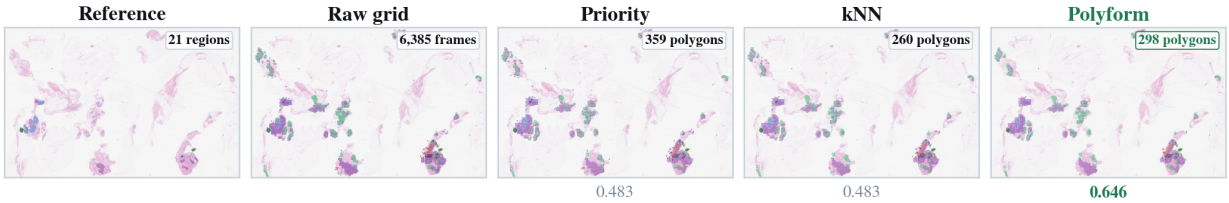

Figure 4. Qualitative comparison of postprocessing strategies for CONCH-LoRA: the reference annotation, the raw grid output, priority, kNN and polyform. The corresponding MDICE scores are shown below each map.

7. Summary

We compared three pathology foundation models, each frozen and LoRA-adapted, against a CNN baseline on a digital pathology task based on TUBR. CONCH with LoRA was strongest at 0.646 MDICE and 0.544 κ, an 11.20% relative improvement over the baseline. Every frozen model exceeded the baseline, and LoRA further improved CONCH and GigaPath. A label-preserving region former, polyform, converted the frame grid into valid, non-overlapping polygons that substantially outperform the conventional relabeling postprocessors.

Future work will include a per-encoder LoRA study across a larger set of foundation models and measuring detection error on densely annotated control slides. The adaptation-and-decoding toolkit is available from the Neural Engineering Data Consortium web site (*www.nedcdata.org*).

Acknowledgments

This material is based on work supported by the National Science Foundation (grant no. 2211841) and the Temple University Office of the Vice-Provost for Research STEM Research Grants. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of these organizations.

References

[1] M.A. Al Mamun, S. Purba, I. Obeid, and J. Picone, "Automated Interpretation and Visualization in Digital Pathology," in *Proceedings of the Northeast Bioengineering Conference*, Philadelphia, Pennsylvania, USA: IEEE, Apr. 2026, pp. 1-3. url: *https://isip.piconepress.com/publications/conference_presentations/2026/nebec/dpath*.

[2] D. Hackel, M. Bagritsevich, C. Dumitrescu, Md. A. Al Mamun, S. A. Purba, D. Heathcote, I. Obeid, and J. Picone, "Enabling Microsegmentation: Digital Pathology Corpora for Advanced Model Development," in Signal Processing in Medicine and Biology: Applications of Artificial Intelligence in Medicine and Biology, New York City, New York, USA: Springer, 2026, p. 50. url: *https://isip.piconepress.com/publications/book_sections/2026/springer/dpath*.

[3] S. S. Shalamzari, M. Bagritsevich, Melles, Anne-Mai, I. Obeid, J. Picone, D. Connolly, C. Wu, B. Brown, J. James, Y. Gong, and H. Wu, "Big Data Resources for Digital Pathology," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium*, Philadelphia, Pennsylvania, USA: IEEE, 2023, pp. 1-19. *doi: 10.1109/SPMB59478. 2023.10372721*.

[4] N. Shawki et al., "The Temple University Digital Pathology Corpus," in *Signal Processing in Medicine and Biology: Emerging Trends in Research and Applications*, 1st ed., I. Obeid, I. Selesnick, and J. Picone, Eds., New York City, New York, USA: Springer, 2020, pp. 67-104. doi: *10.1007/978-3-030-36844-9*.

[5] B. Ehteshami Bejnordi 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, Dec. 2017, doi: *10.1001/jama.2017.14585*.

[6] G. Campanella et al., "Clinical-grade computational pathology using weakly supervised deep learning on whole slide images," *Nature Medicine*, 2019, Vol.25 (8), pp. 1301-1309, 2019, doi: *10.1038/s41591-019-0508-1*.

[7] M. Y. Lu, D. F. K. Williamson, T. Y. Chen, R. J. Chen, M. Barbieri, and F. Mahmood, "Data-efficient and weakly supervised computational pathology on whole-slide images," *Nature Biomedical Engineering*, vol. 5, no. 6, pp. 555-570, Jun. 2021, doi: *10.1038/s41551-020-00682-w*.

[8] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proceedings of the International Conference on Learning Representations* (ICLR), Vienna, Austria: OpenReview.net, 2021, pp. 1–21. doi: *https://iclr.cc/virtual/2021/oral/3458.*

[9] M. Oquab et al., "DINOv2: Learning Robust Visual Features without Supervision.," Trans. Mach. Learn. Res., vol. 2024, 2024, url: *https://openreview.net/forum?id=a68SUt6zFt*.

[10] R. J. Chen et al., "Towards a general-purpose foundation model for computational pathology.," *Nat Med*, vol. 30, no. 3, pp. 850–862, Mar. 2024, doi: *10.1038/s41591-024-02857-3*.

[11] M. Y. Lu et al., "A visual-language foundation model for computational pathology.," Nat Med, vol. 30, no. 3, pp. 863–874, Mar. 2024, doi: *10.1038/s41591-024-02856-4*.

[12] H. Xu et al., "A whole-slide foundation model for digital pathology from real-world data," *Nature*, vol. 630, no. 8015, pp. 181–188, Jun. 2024, doi: *10.1038/s41586-024-07441-w*.

[13] E. Vorontsov et al., "A foundation model for clinical-grade computational pathology and rare cancers detection," *Nature Medicine*, vol. 30, no. 10, pp. 2924–2935, Oct. 2024, doi: *10.1038/s41591-024-03141-0*.

[14] X. Wang et al., "Transformer-based unsupervised contrastive learning for histopathological image classification," *Medical Image Analysis*, vol. 81, p. 102559, Oct. 2022, doi: *10.1016/j.media.2022.102559*.

[15] E. Hu et al., "LoRA: Low-Rank Adapation of Large Language Models," International Conference on Learning Representations, 2022, doi: *10.48550/arXiv.2106.09685*.

[16] N. Houlsby et al., "Parameter-Efficient Transfer Learning for NLP," in *Proceedings of the 36th International Conference on Machine Learning*, K. Chaudhuri and R. Salakhutdinov, Eds., in Proceedings of Machine Learning Research, vol. 97. PMLR, Jun. 2019, pp. 2790–2799. url: *https://proceedings.mlr.press/v97/houlsby19a.html*.

[17] S. A. Purba, A.-M. Melles, D. Hackel, I. Obeid, and J. Picone, "Assessing the Visual Reasoning of Multimodal Language Models in Biomedical Applications," in *Proceedings of the IEEE Signal Processing in Medicine and Biology Symposium* (SPMB), Philadelphia, Pennsylvania, USA: IEEE, Aug. 2025, pp. 1–8. doi: *10.1109/SPMB67169.2025.11345419.*

[18] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in Proceedings of the International Conference on Machine Learning (ICML), K. Chaudhuri and R. Salakhutdinov, Eds., in *Proceedings of Machine Learning Research*, vol. 97. Long Beach, California, USA: PMLR, May 2019, pp. 6105–6114. url: *http://proceedings.mlr.press/v97/tan19a.html*.

[19] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (CVPR), Las Vegas, Nevada, USA, 2016, pp. 770–778. doi: *10.1109/CVPR.2016.90.*

[20] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* (MICCAI), N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds., Munich, Germany: Springer International Publishing, 2015, pp. 234–241. doi: *10.1007/978-3-319-24574-4_28*.

[21] S. Graham et al., "Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images.," *Med Image Anal*, vol. 58, p. 101563, Dec.2019, doi: *10.1016/j.media.2019.101563*.

[22] N. Otsu, "A Threshold Selection Method from Gray-Level Histograms," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979, doi: *10.1109/TSMC.1979.4310076*.

[23] C. Dumitrescu et al., "Rapid and Inexpensive Precision Breast Cancer Screening Using Machine Learning," Pennsylvania Breast Cancer Consortium, Harrisburg, Pennsylvania, USA, Aug. 2025. url: *https://isip.piconepress.com/publications/reports/2025/pabcc/dpath/*.

[24] J. Cohen, "A Coefficient of Agreement for Nominal Scales," *Educational and Psychological Measurement*, vol. 20, pp. 37–46, 1960. doi: *10.1177/0013164491511008*.

[25] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization," in *Proceedings of the International Conference on Learning Representations*, 2019. doi: *10.48550/arXiv.1711.05101*.