LiNA: Rethinking Shape Capacity for Efficient Nuclei Analysis

Anonymous authors
Paper under double-blind review

Abstract

How much shape-decoding capacity does joint nuclei analysis actually need? High-performing systems devote substantial computation to predicting instance geometry, yet it is unclear whether this cost is justified by the complexity of nuclear shape. If nuclear geometry needs only a few degrees of freedom, dense shape decoding may spend much of that computation on capacity with little task-level benefit. We probe this by varying Fourier order N, which controls the boundary's degrees of freedom. Across four histopathology datasets, a single Fourier harmonic reconstructs ground-truth contours at 85–89% IoU and order $N{=}15$ reaches 98.5–99.1%, with an independent boundary-distance measure showing the same low-order saturation. These results raise the possibility that current pipelines devote more capacity to nuclear geometry than the task requires. To test and exploit this observation, we instantiate LiNA (Lightweight Nuclei Analysis), which separates localization and typing from geometry prediction, represents each nucleus with a compact Fourier contour, and strengthens typing through prototype distillation. We then replace its Fourier pathway with a matched dense-mask decoder while keeping the remaining system fixed. Across four backbone scales, the dense-mask decoder improves none of the evaluated task metrics, while computation grows with nucleus density and reaches $6.4\times$ that of the matched Fourier variant on the densest PanNuke tile for ConvNeXt-Small. This compact geometry pathway is nevertheless sufficient for strong joint analysis. At 57M parameters, roughly 8% of 700M-scale SAM-based pipelines, LiNA achieves the best performance on PanNuke and Lizard and the strongest zero-shot results on MoNuSeg and CPM17. Together, these results reveal a shape-capacity mismatch in joint nuclei analysis. LiNA exploits this mismatch to reduce geometry capacity without sacrificing recognition and, combined with its efficient overall design, reaches a new Pareto-efficient operating point among the compared models.

1 Introduction

Nuclei detection, classification, and segmentation are fundamental to computational pathology, as the count, type distribution, and morphology of nuclei inform diagnosis, grading, and prognosis. Recent joint-analysis systems increasingly use larger models and incur greater inference cost to achieve higher accuracy. Performance-oriented methods often rely on powerful encoders and dense or multi-stage decoding, whereas lightweight alternatives typically reduce computational cost at some expense to recognition accuracy. As summarized in Fig. 2a, existing methods therefore exhibit a clear accuracy–efficiency trade-off. However, this trade-off has largely been observed under designs that scale capacity up or down across the system, and it remains unclear whether the same trade-off is unavoidable when the capacity of a particular prediction pathway is adjusted selectively. We therefore examine model capacity at the component level. Focusing on shape decoding, we ask a more specific question. How much shape-decoding capacity does nuclei analysis actually require?

To answer this question, we define shape capacity as the capacity devoted to representing and predicting instance geometry, and analyze it directly. We first ask, independently of any specific model architecture, how much representational complexity nuclear geometry itself demands. Across four histopathology datasets, when Fourier contours fitted to ground-truth mask boundaries are rasterized back to masks, a single Fourier harmonic already reaches 85–89% IoU with the originals, while $N{=}15$ reaches 98.5–99.1%. An independent boundary-distance metric shows the same low-order

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

knee with sharply diminishing returns (Fig. 4). These results suggest that nuclear geometry may require relatively little representational capacity to reach high fidelity, raising the possibility that current systems devote more capacity to geometry prediction than the task actually requires.

To test and exploit this possibility in a complete joint-analysis system, we instantiate LiNA (Lightweight Nuclei Analysis). LiNA represents each nucleus with a compact low-order Fourier contour rather than a dense mask. Compact Fourier contours have also been used in prior biomedical instance-segmentation methods Upschulte et al. (2022); Hirling & Horvath (2023), which established their effectiveness as compact representations for instance segmentation. However, these methods primarily target binary instance segmentation and do not analyze how much geometry capacity joint analysis actually requires when detection and fine-grained nucleus typing are also involved. LiNA separates point-based localization and instance-level typing from geometry prediction. Detection and typing are obtained directly from instance-level features rather than from the geometry output, so the geometry representation's capacity can be varied without changing how detection and typing are computed. A training-time prototype distillation pathway further strengthens fine-grained typing.

We then vary shape capacity directly within LiNA to test this hypothesis. As Fourier order increases, segmentation quality rises sharply at low orders and saturates around $N{=}10$, while detection and classification remain stable across the sweep. For a more direct test, we replace the Fourier pathway with a matched per-instance dense-mask decoder while keeping the rest of the system and training recipe fixed. Across four backbone scales, the dense-mask decoder improves none of the evaluated task metrics. Its computation, however, grows sharply with nucleus density and reaches up to $6.4\times$ that of the matched Fourier variant for ConvNeXt-Small.

This compact geometry pathway is nevertheless sufficient for strong joint-analysis performance. On PanNuke, LiNA achieves 0.88/0.72 detection/classification F1 with approximately 57M parameters and 48G FLOPs, outperforming much larger SAM-based joint-analysis pipelines in detection and classification. On Lizard, it achieves 0.90/0.71 detection/classification F1 and 0.46 avgPQ. When trained on PanNuke and transferred without fine-tuning, LiNA also achieves the strongest zero-shot results among the compared methods on MoNuSeg and CPM17. Together, these results reveal a shape-capacity mismatch in joint nuclei analysis. Nuclear geometry reaches high fidelity with a compact representation, while more expensive dense shape decoding provides no measurable task-level benefit in our controlled comparisons. By exploiting this mismatch, LiNA selectively reduces the capacity devoted to geometry while preserving recognition performance and, together with its efficient overall design, establishes a new Pareto-efficient operating point among the compared joint-analysis systems (Fig. 2a).

We make three contributions.

- We identify a shape-capacity mismatch in joint nuclei analysis. Across four histopathology datasets, nuclear geometry exhibits low representational complexity, while a matched dense-mask decoder requires substantially more computation without improving any evaluated task metric.
- We instantiate this finding in LiNA, a complete joint-analysis system. LiNA combines a compact instance-geometry representation with point-based localization and instance-level typing, while prototype distillation strengthens fine-grained classification for efficient joint detection, classification, and segmentation.
- We demonstrate strong accuracy and efficiency together with robust cross-dataset generalization across PanNuke, Lizard, MoNuSeg, and CPM17. LiNA achieves the highest detection and classification performance among the compared methods on PanNuke, remains strong on Lizard and cross-dataset evaluation, and establishes a new Pareto-efficient operating point among the compared joint-analysis systems.

2 Related Work

2.1 Performance-Oriented Nuclei Analysis.

Performance-oriented systems increasingly combine large pretrained encoders with dense or prompt-based decoders Jeong et al. (2026); Hörst et al. (2026). CellViT Hörst et al. (2024) uses

SAM with HoVer-Net-style dense prediction and watershed post-processing, whereas SAC Na et al. (2025) and PromptNucSeg Shui et al. (2024a) adopt SAM-based prompting in a two-stage pipeline. These systems achieve strong accuracy but incur substantially higher parameter, computation, and latency costs, and the prompt-based two-stage designs further add the cost of a separate prompt-generation stage before mask prediction. In real-world pathology workflows, deployment environments vary in available compute, while slide-level throughput remains a practical constraint, making such computational costs increasingly important at scale. This motivates treating deployment cost as an explicit design constraint in nuclei analysis rather than as a secondary consideration.

2.2 Efficiency-Oriented Nuclei Analysis.

To address this computational cost problem and make nuclei analysis deployable under limited resources, a variety of lightweight methods have been proposed in recent years. NuLite Tommasino et al. (2026) adopts a FastViT-based lightweight U-Net architecture, substantially reducing parameter count and computation. FKDNuSeg Zhao et al. (2026) combines an ENet backbone with a multi-task learning structure under knowledge distillation, performing nuclei segmentation and classification with over 10× fewer parameters than prior SOTA models. These methods substantially reduce computation, but their reported results still expose an accuracy–efficiency trade-off. A complementary route reduces the complexity of the output representation itself. StarDist Schmidt et al. (2018) models each instance as a star-convex polygon, CPP-Net Chen et al. (2023b) strengthens polygonal instance modeling with contextual prediction, Contour Proposal Networks (CPN) Upschulte et al. (2022) regress a fixed-size Fourier-descriptor contour per instance in a single forward pass, and FourierNet Hirling & Horvath (2023) predicts Fourier-based shape priors in a dense per-location formulation. Collectively, these methods establish that compact shape parameterizations are effective for biomedical instance segmentation. Their formulations, however, target binary instance segmentation and therefore leave open the central question for multi-class joint analysis, namely how to reduce the cost of instance geometry while preserving strong localization and fine-grained nucleus typing. We provide detailed controlled comparisons with these prior Fourier-descriptor methods in Appendix C, under both their native binary task and a diagnostic multiclass extension.

3 Method

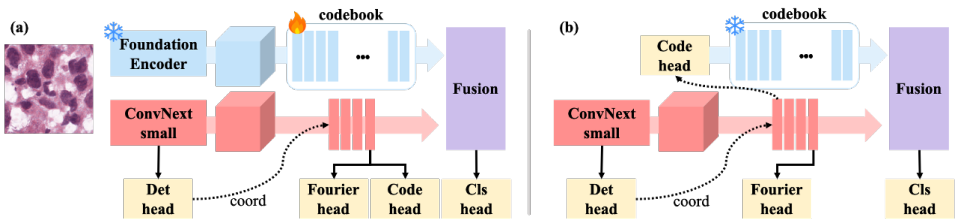


Figure 1: Overview of the proposed framework. (a) and (b) illustrate the training and the inference stages, respectively.

To instantiate this finding in joint nuclei analysis, LiNA uses a one-stage end-to-end architecture that performs nuclei detection, classification, and segmentation in a single forward pass (Fig. 1). Given an input image, a student encoder extracts multi-scale feature maps. A detection head then localizes nuclei as points, placing a fixed set of center candidates on a regular grid and regressing their offsets to nucleus centers, following deformable point-based detection Shui et al. (2024b); Song et al. (2021). A light NMS then removes duplicate points, and for each predicted center, we sample a region-of-interest (ROI) feature from the student feature maps by bilinear interpolation. We feed it to two parallel branches, a segmentation branch that regresses Fourier coefficients into an instance contour and a classification branch that predicts the nucleus type.

LiNA separates recognition from shape prediction. Detection and typing are read out from point-wise features, while segmentation is represented by a compact Fourier contour, avoiding dense upsampling and watershed-based instance recovery. Prototype distillation into a compact codebook Linde et al. (1980) supplies richer semantic supervision during training, while inference runs in a single forward pass and remains teacher-free.

3.1 CONTOUR PREDICTION FOR EFFICIENT SEGMENTATION.

Most nuclei instance segmentation approaches predict a pixel-wise mask and rely on heavy upsampling decoders to recover full-resolution masks from compressed backbone features. This dense decoding is computationally expensive and often needs additional post-processing to separate adjacent nuclei, which is at odds with our compact single-stage design. We therefore describe each nucleus boundary as a closed contour. As shown in Fig. 1, the detection head predicts each nucleus center, and the feature at that predicted coordinate is sampled from the student feature map and passed to the Fourier head as input. The head then regresses a compact set of Fourier coefficients, which reconstruct the contour without any pixel-level mask decoding. This contour is supervised by matching its points to the ground-truth contour points, with the full loss detailed in the loss function.

Concretely, we treat a nucleus boundary as a closed curve in the complex plane. Sampling the contour at parameter $t \in [0, 1)$ and writing each point as $z(t) = x(t) + i\, y(t)$, we represent it by a truncated Fourier series Zhu et al. (2021)

$$\hat{z}(t) = \sum_{k=-N}^{N} c_k\, e^{i2\pi kt}, \qquad c_k \in \mathbb{C}, \tag{1}$$

where the $2N{+}1$ complex coefficients $\{c_k\}$ are predicted from the ROI feature at the nucleus center, and c_0 encodes the contour centroid as a correction to the predicted center. In practice, the head is a lightweight MLP that outputs the $2(2N{+}1)$ real and imaginary parts, and the contour is reconstructed by evaluating Eq. 1 at sampled t. Because reconstruction is a direct sum of complex exponentials, it needs no spatial upsampling, keeping the segmentation head extremely light and completing detection, classification, and segmentation in a single feed-forward pass.

Beyond efficiency, the Fourier contour is structurally convenient under extreme model compression. Each detected nucleus is represented by one closed contour by construction, and each detected center carries its own instance identity. This removes the need for a separate connected-component or watershed step to recover instances from a shared dense mask, because instances are never merged in a shared mask to begin with.

3.2 PROTOTYPE DISTILLATION AND FUSION.

The classification branch faces a challenge under a lightweight backbone. A compact student encoder learns data-specific representations that are useful on their own but do not carry the robust, general representations of a pretrained foundation model. Since the two are different and complementary, combining them lets the model draw on the strengths of both, and this benefit is most pronounced for under-represented classes where the student alone is weak. Running the teacher at inference, however, would eliminate the very efficiency we target. We therefore introduce a two-phase Prototype Distillation (proto-distill) that transfers the teacher representation into a compact codebook, which the student consults at inference without the teacher. Unlike prior VQ-based distillation that reconstructs teacher features Peng et al. (2022) or predicts hard codebook indices Guo et al. (2023), we align a soft code distribution with a KL objective and fuse it into classification, per nucleus over foreground regions.

Phase 1: Codebook Learning. This phase learns the prototype codebook that distills the frozen teacher's semantics. At each predicted center coordinate $\mathbf{c}$, we sample the frozen teacher encoder Filiot et al. (2025) feature via bilinear grid sampling. The sampled features are projected to a prototype dimension D via a projection head f_{proj}, and then attend over a learnable Prototype Codebook $\mathbf{E} \in \mathbb{R}^{K\times D}$ via cosine-similarity-based soft attention with temperature τ:

$$\mathbf{z}_i^{\text{t}} = \text{softmax}\left(\frac{\tilde{\mathbf{f}}_i^{\text{t}} \cdot \tilde{\mathbf{E}}^{\top}}{\tau}\right) \in \mathbb{R}^{K}, \tag{2}$$

where $\tilde{(\cdot)}$ denotes ℓ_2-normalization. The teacher prototype embedding is computed as $\mathbf{e}_i^{\text{t}} = \mathbf{z}_i^{\text{t}}\, \mathbf{E}$. For nuclei classification, we fuse the student ROI feature $\mathbf{r}_i$ with the prototype embedding by concatenation, i.e., $\hat{\mathbf{f}}_i = [\mathbf{r}_i \parallel \mathbf{e}_i^{\text{t}}]$, and feed $\hat{\mathbf{f}}_i$ to the classification head.

Foreground-Only Constraint. Codes are computed at every grid location, but pathology images are dominated by large, uninformative background regions, so we learn the codebook only at foreground

(nuclei) grid locations. We downsample the ground-truth binary foreground mask $\mathbf{m}$ to the grid resolution to select these locations, and learn the codebook and apply the distillation loss only there, so the codebook encodes only pathologically relevant semantics.

Phase 2: Student Code Predictor Distillation. This phase learns a student code predictor that reproduces the teacher codes from the student's own features. Once Phase 1 converges, the teacher projection head and codebook $\mathbf{E}$ are fully frozen. The lightweight Code Predictor g_{pred} maps a student ROI feature to a soft distribution over the K prototypes, the same mixing weights the teacher produces. Training pulls this student distribution toward the teacher's, so the student reproduces the teacher's prototype mixture from its own features alone. The distillation objective is the foreground-masked KL divergence Kullback & Leibler (1951) between teacher codes $\mathbf{z}^{\text{t}}$ and student codes $\mathbf{z}^{\text{s}} = \text{softmax}(g_{\text{pred}}(\mathbf{r}) \,/\, \tau)$:

$$\mathcal{L}_{\text{proto}} = \frac{1}{|\mathcal{F}|} \sum_{i \in \mathcal{F}} D_{\text{KL}}\big(\mathbf{z}_i^{\text{t}} \parallel \mathbf{z}_i^{\text{s}}\big)\,, \tag{3}$$

where $\mathcal{F}$ denotes the set of foreground grid locations, obtained by downsampling the GT mask to the grid resolution.

Teacher-Free Inference. Inference is fully teacher-free. The student code predictor estimates the code distribution $\mathbf{z}^{\text{s}}$ directly from its ROI feature, the corresponding prototype embedding $\mathbf{z}^{\text{s}}\mathbf{E}$ is looked up from the frozen codebook and concatenated with the student feature for classification, and the teacher encoder is never invoked.

3.3 Loss Function and Optimization.

The framework is trained end-to-end with a total loss that is a weighted sum of five complementary terms, $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{det}} + \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{fourier}} + \mathcal{L}_{\text{CAB}} + \mathcal{L}_{\text{proto}}$. After establishing optimal prediction-to-GT correspondences via Hungarian matching, $\mathcal{L}_{\text{det}}$ applies Smooth L1 loss on matched cell centroid coordinates. $\mathcal{L}_{\text{cls}}$ is a multi-label binary cross-entropy loss that classifies the pathological cell type of each matched proposal. For cell shape, rather than comparing coefficients directly, $\mathcal{L}_{\text{fourier}}$ decodes both predicted and GT Fourier coefficients into 128-point contours and matches those points with a shift-invariant Smooth L1 loss Zhu et al. (2021), resolving the start-point ambiguity inherent in closed contour prediction. This contour matching constrains the global shape but leaves it only loosely tied to the pixel-level boundary, and a low-order series is biased toward smooth, ellipse-like outlines that blur fine detail such as sharp corners. $\mathcal{L}_{\text{CAB}}$, our Curvature-Aware Boundary loss, addresses both issues by aligning the contour with the true cell boundary,

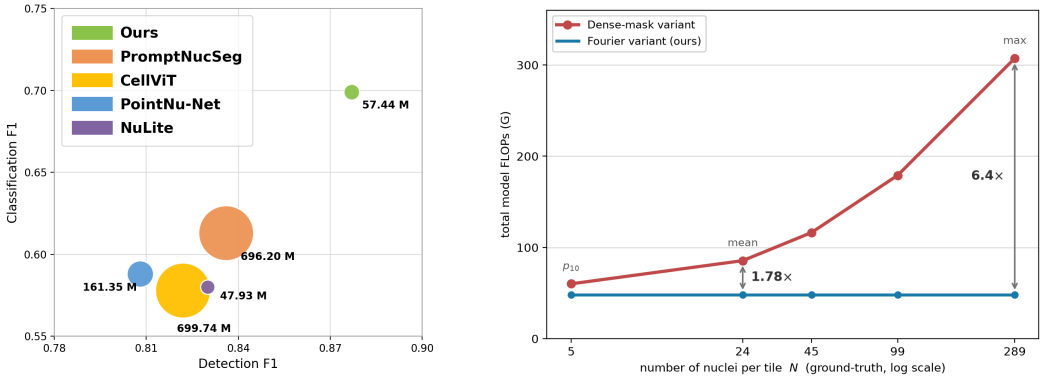
$$\mathcal{L}_{\text{CAB}} = \frac{1}{|\mathcal{C}|} \sum_{c \in \mathcal{C}} \rho\big(d(c, \mathcal{B})\big) + \frac{\beta}{|\mathcal{P}|} \sum_{p \in \mathcal{P}} \rho\big(d(p, \mathcal{C})\big), \tag{4}$$

where $d(x, \mathcal{S}) = \min_{s \in \mathcal{S}} \|x - s\|_2$ is the distance from a point to a set, ρ is the Smooth L1 penalty toward zero, $\mathcal{B}$ is the GT cell boundary, $\mathcal{C}$ the points decoded from the predicted Fourier contour, $\mathcal{P} \subset \mathcal{B}$ the high-curvature landmark points, and β balances the two terms. The first term pulls the predicted contour onto the boundary, where $d(c, \mathcal{B})$ is the absolute value of a precomputed signed distance field of the GT mask, sampled at the contour points for efficiency. The second pulls the boundary's high-curvature corners onto the contour to preserve the sharp detail a low-order representation tends to smooth away. Cell centroids are detached so that this loss supervises only the Fourier coefficients, not centroid prediction. Finally, $\mathcal{L}_{\text{proto}}$ is the foreground-aware KL divergence defined in Eq. 3, activated exclusively in Phase 2. All terms except $\mathcal{L}_{\text{proto}}$ are active from the start, and the prototype codebook is learned through the classification loss $\mathcal{L}_{\text{cls}}$ in Phase 1.

4 Experiments

4.1 Dataset.

PanNuke has 7,899 images of 256×256 from 19 tissue types, and we follow its official cross-validation protocol Gamper et al. (2020) and report our main results averaged over the three folds. **MoNuSeg** has 44 images of 1,000×1,000 across 9 organs (28,846 nuclei). **Lizard** is a colon histology dataset with 495,000 nuclei in six categories. **CPM17** has 64 multi-organ cancer images at 500×500 or 600×600.



(a) Detection vs. classification F1.

(b) Total FLOPs vs. number of nuclei per tile.

Figure 2: Efficiency overview. (a) LiNA leads detection and classification with relatively small size, and circle size denotes parameter count. (b) Replacing the Fourier contour with a per-instance dense mask decoder makes total FLOPs grow with the number of nuclei per tile, up to $6.4\times$ the matched Fourier variant on the densest PanNuke tiles (ConvNeXt-Small), without improving any evaluated task metric.

4.2 Implementation Details.

The student backbone is a pretrained DINOv3 Siméoni et al. (2025) ConvNext Small Woo et al. (2023). We use Fourier degree $N = 15$ (decoded into 128 points) and a prototype codebook of $K = 128$ prototypes of dimension $D = 256$. Training uses AdamW Loshchilov & Hutter (2019) (lr=1×10^{-4}, batch 80) on a single A6000 48GB GPU (PyTorch 2.6, CUDA 12.4, Ubuntu 24.04), with the teacher H0-mini (86M parameters) Filiot et al. (2025) frozen throughout. Phase 2 starts at epoch 70 and the CAB loss at epoch 10. The $\mathcal{L}_{\text{fourier}}$, $\mathcal{L}_{\text{CAB}}$, and $\mathcal{L}_{\text{proto}}$ weights are 0.5, 0.5, and 2.0. The remaining hyperparameters follow Shui et al. (2024a), with all final values and search ranges in the released configs and the supplementary material. Zero-shot models use the same PanNuke split Hörst et al. (2024), and we follow the evaluation protocol of PromptNucSeg Shui et al. (2024a).

4.3 Main Results.

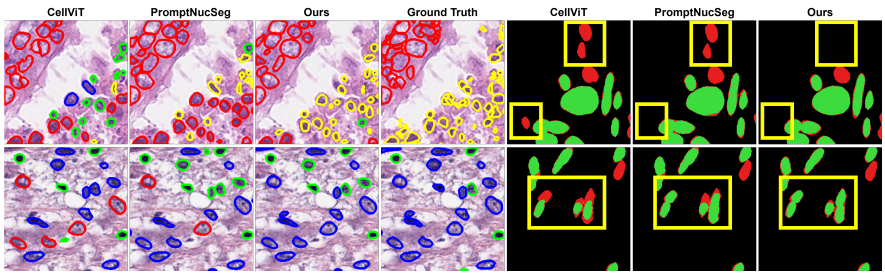


Figure 3: Qualitative results. Left: model predictions. Right: pixel-level correctness, correct pixels in green and errors in red.

On PanNuke, LiNA achieves the best detection and classification among the compared methods (0.88 / 0.72 F1) and reaches panoptic quality on par with the best (Tables 1–2). Detection and classification are read out independently of the segmentation representation, and the controlled head-swap shows that substantially increasing shape-decoding capacity does not improve either task (Table 5b). The fused prototype codes further sharpen classification (Table 6). Crucially, its single-stage end-to-end design and a lightweight Fourier contour head let LiNA attain this accuracy at an order of magnitude fewer parameters and several times lower latency than 700M-scale pipelines such as CellViT and PromptNucSeg (Table 3). It therefore occupies a stronger accuracy–efficiency operating point than prior joint-analysis systems (Fig. 2a). Qualitative results are shown in Fig. 3. On Lizard,

Table 1: Detection and classification F1 on PanNuke. Predictions are matched to the nearest ground-truth nucleus within $\approx 3\mu$m (class-agnostic for detection, same-class for classification). Best and second-best in **bold** and underline.

Method	Detection	Classification					
	Avg	Neo	Epi	Inf	Con	Dead	Avg
Hover-Net Graham et al. (2019)	0.80	0.62	0.56	0.54	0.49	0.31	0.50
StarDist Schmidt et al. (2018)	0.82	0.69	0.70	0.57	0.51	0.10	0.52
CP-Mamba Zhang et al. (2025)	0.82	0.62	0.64	0.54	0.50	0.31	0.52
CPP-Net Chen et al. (2023b)	0.82	0.70	0.72	0.58	0.53	0.38	0.58
NuLite-H Tommasino et al. (2026)	0.83	0.71	0.73	0.58	0.53	0.37	0.58
HistoNeXt Chen et al. (2025)	0.82	0.70	0.71	0.58	0.52	0.36	0.57
PointNu-Net Yao et al. (2024)	0.81	0.73	0.73	<u>0.60</u>	<u>0.58</u>	0.31	0.59
CellViT Hörst et al. (2024)	0.83	0.71	0.73	0.58	0.53	0.36	0.58
MOTDNet Yang et al. (2026)	0.83	0.70	0.72	0.59	0.52	0.36	0.58
PromptNucSeg Shui et al. (2024a)	<u>0.84</u>	<u>0.71</u>	<u>0.76</u>	0.59	0.55	<u>0.46</u>	<u>0.61</u>
Ours	**0.88**	**0.81**	**0.83**	**0.74**	**0.68**	**0.54**	**0.72**

Table 2: Per-class panoptic quality (PQ) on PanNuke. *Average* is the mean over the five classes (avgPQ). The best scores are shown in **bold**.

Method	Neoplastic	Epithelial	Inflammatory	Connective	Dead	Average
FKDNuSeg	0.551	0.483	0.408	0.367	0.150	0.392
HoVer-Net	0.551	0.491	0.417	0.388	0.139	0.397
StarDist	0.547	0.532	0.424	0.380	0.123	0.401
CPP-Net	0.571	0.565	0.405	0.395	0.131	0.413
NuLite-H	0.577	0.571	0.417	0.413	0.145	0.425
MOTDNet	0.578	0.563	0.440	0.410	0.136	0.425
PointNu-Net	0.578	0.577	0.433	0.409	0.154	0.430
CellViT	0.581	<u>0.583</u>	0.417	0.423	0.149	0.431
PromptNucSeg	<u>0.598</u>	0.582	<u>0.441</u>	**0.433**	<u>0.161</u>	<u>0.443</u>
Ours	**0.599**	**0.598**	**0.445**	**0.433**	**0.180**	**0.451**

Table 3: Efficiency comparison on PanNuke (Params in M, FLOPs in G, latency in ms). Ours-T uses a ConvNeXt-Tiny backbone.

Method	Params	FLOPs	Lat.	Det	Cls	avgPQ
FKDNuSeg	4.1	21.6	–	–	–	0.392
NuLite-T	12.05	158.52	17.13	0.82	0.57	0.419
NuLite-H	47.93	265.35	32.75	0.83	0.58	0.425
MOTDNet	32.12	13.36	–	0.83	0.58	0.425
PointNu-Net	161.35	560.30	58.53	0.81	0.59	0.430
CellViT	699.74	942.54	64.00	0.83	0.58	0.431
PromptNucSeg	696.20	570.44	63.83	0.84	0.61	0.443
Ours-T	36.36	36.93	9.9	<u>0.88</u> $\pm$ 0.002	<u>0.71</u> $\pm$ 0.015	<u>0.449</u> $\pm$ 0.010
Ours	57.44	47.99	11.55	**0.88** $\pm$ 0.002	**0.72** $\pm$ 0.015	**0.451** $\pm$ 0.008

scored with a single identical scorer under a matched $40\times$ protocol (Table 4), LiNA again leads detection and classification (0.90 / 0.71 F1) achieving the highest avgPQ in the table while remaining relatively small and the fastest model in the table. The same accuracy–efficiency pattern on a second benchmark shows that the result extends beyond PanNuke.

Shape-Capacity Saturation. Reconstruction fidelity saturates at low representation order. Fitting a truncated Fourier series to ground-truth contours across four datasets, a single harmonic already reaches 85–89% IoU, and the order $N{=}15$ used by LiNA reaches 98.5–99.1% (median 98.9–99.3%),

Table 4: Results on the Lizard dataset at matched $40\times$ magnification (512×512). As PromptNucSeg is two-stage, its Params/FLOPs/Latency are summed over both stages.

Method	Params	FLOPs	Lat.	Det	Cls	avgPQ	bPQ
CellViT (ViT256)	46.74	1108.20	52	0.87	0.66	0.42	0.64
CellViT (SAM-B)	146.09	1803.80	75	0.88	0.68	0.43	0.65
CellViT (SAM-L)	367.84	2396.88	118	0.88	0.68	0.44	0.65
CellViT (SAM-H)	699.74	3266.66	173	0.88	0.70	0.45	0.65
PromptNucSeg (SAM-B)	148.85	409.26	75	0.85	0.66	0.43	0.64
PromptNucSeg (SAM-H)	696.20	1769.04	178	0.85	0.66	0.44	0.65
Ours	57.44	191.75	16	**0.90**	**0.71**	**0.46**	**0.66**

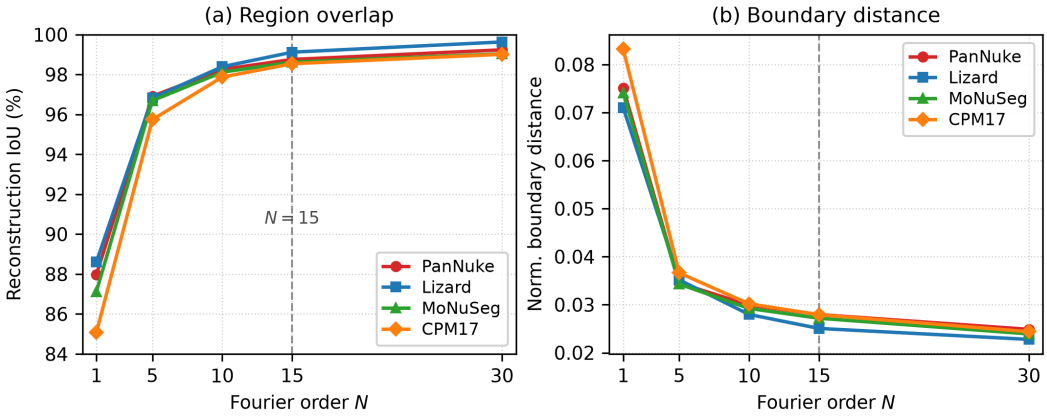


Figure 4: Fidelity of an order-N Fourier fit to ground-truth nucleus contours (centroid-aligned) on four datasets, as (a) region overlap (IoU) and (b) normalized boundary distance. Both show a clear knee by $N \approx 10$–15 (LiNA uses N=15) with diminishing returns beyond. Protocol in Appendix E.

with about half a percentage point of further gain at N=30 (Fig. 4a). At N=15, 99.6–100% of nuclei exceed 0.90 IoU and 97.5–99.9% exceed 0.95, even in the lowest-solidity quartile ($\geq$97.8% mean). An independent normalized boundary-distance metric shows the same low-order knee and strongly diminishing returns (Fig. 4b), so the effect is not an artifact of region-overlap scoring. A low-order contour thus captures nuclear geometry at high fidelity, leaving little additional fidelity for a substantially higher-capacity representation to recover.

We next examine shape capacity in the trained system along two axes: Fourier representation order and decoding pathway. Varying the Fourier order reveals a clear low-capacity regime at N=1: segmentation quality rises steeply at the lowest orders and saturates by N=10, with only negligible changes in avgPQ and bPQ beyond (Table 5a). Detection and classification remain essentially unchanged across the entire sweep. We then train matched Fourier and dense variants end-to-end under an identical recipe, changing only the shape pathway. Replacing the Fourier contour with a per-instance dense mask decoder improves no evaluated task metric across four backbone scales, yet total FLOPs diverge sharply (Table 5b). A dense decoder produces one mask per nucleus, so its cost grows with the number of nuclei per tile, reaching up to $6.4\times$ the matched Fourier variant on the densest PanNuke tiles for ConvNeXt-Small, whereas the analytic Fourier contour is nearly free per instance (Fig. 2b). A compact contour representation therefore keeps the geometry-decoding cost nearly flat as nucleus density grows.

Ablation Study. Table 6 shows that CAB raises avgPQ from 0.413 to 0.436, while proto-distill and fusion further improve cls-F1 to 0.719 and avgPQ to 0.451. Randomized codes and removal of the foreground constraint reduce performance, with the largest class-wise gain on the rare Dead class (Table 7). Proto-distill also outperforms feature- and logit-level distillation.

Table 5: Shape-capacity interventions on PanNuke. (a) Fourier-order sweep (ConvNeXt-S, 3-fold mean; Δ from the preceding order). (b) Matched Fourier (-F) and dense-mask (-D) pathways across backbones. F_μ and $F_{\max}$ are total FLOPs (G) at the mean and maximum observed nucleus densities.

(a) Fourier order N (Small)

	N=1	N=3	N=5	N=10	N=15
Det	0.88	0.88	0.88	0.88	0.88
Cls	0.72	0.72	0.72	0.72	0.72
avgPQ	0.438	0.448	0.450	0.451	0.451
bPQ	0.660	0.674	0.678	0.679	0.679
ΔavgPQ	–	+0.010	+0.002	+0.001	+0.000
ΔbPQ	–	+0.014	+0.004	+0.001	+0.000
Params	57.43	57.43	57.43	57.43	57.44
FLOPs	47.98	47.98	47.98	47.98	47.99

(b) Fourier vs. dense mask across scales

Backbone	Det/Cls	avgPQ	F_μ	$F_{\max}$
Tiny-F	0.88/0.71	0.44	36.9	36.9
Tiny-D	0.87/0.70	0.44	74.5	296.3
Small-F	0.88/0.72	0.45	48.0	48.0
Small-D	0.87/0.70	0.44	85.6	307.4
Base-F	0.88/0.72	0.45	65.8	65.8
Base-D	0.88/0.71	0.44	103.4	325.2
Large-F	0.88/0.72	0.45	116.2	116.2
Large-D	0.88/0.71	0.45	153.8	375.6

Table 6: Ablation of key components on the PanNuke dataset. The indented rows ablate proto-distill.

Components	cls-F1	avgPQ
Baseline	0.622	0.413
+ CAB loss	0.640	0.436
+ Proto-Distill and Fusion	0.719	0.451
w/o foreground mask	0.681	0.353
code head random	0.673	0.435
Feature Distill	0.683	0.424
Logit Distill	0.684	0.426

Table 7: Per-class classification F1 for different code selection strategies (ConvNeXt-T).

Selection	Neo	Epi	Inf	Conn	Dead	All
Proto-Distill	80.8	83.1	73.8	67.6	**47.6**	70.6
Random	79.8	81.0	73.3	67.2	33.1	66.9

Table 8: Performance on MoNuSeg and CPM17. Our model is trained on PanNuke without any fine-tuning.

Method	MoNuSeg		CPM17	
	AJI	PQ	AJI	PQ
Supervised				
HARU-Net	0.613	0.572	0.721	0.701
Hover-Net	0.618	0.597	0.705	0.697
Triple U-Net	0.621	0.601	0.711	0.685
PointNu-Net	0.606	0.603	0.712	0.706
FEEDNet	0.616	0.613	0.701	0.705
PromptNucSeg	0.622	0.627	0.740	0.733
Zero-shot (no fine-tuning)				
PromptNucSeg	0.532	0.559	0.625	0.666
NuLite-T	0.628	0.630	0.670	0.679
NuLite-H	0.627	0.636	0.667	0.681
CellViT	0.628	0.639	0.691	0.694
Ours	**0.630**	**0.644**	**0.700**	**0.706**

Cross-Dataset Generalization. Without fine-tuning, LiNA achieves the strongest zero-shot results among the compared methods on MoNuSeg (0.630/0.644 AJI/PQ) and CPM17 (0.700/0.706), while remaining competitive with fully supervised baselines Chen et al. (2023a); Zhao et al. (2020); Deshmukh et al. (2022); Shui et al. (2024a) (Table 8).

5 Conclusion

Our results challenge the assumption that joint nuclei analysis requires high-capacity shape decoding. Across four datasets, nuclear geometry reaches high fidelity at low representation order, while a matched dense-mask pathway adds substantial computation without improving any evaluated metric. LiNA exploits this shape-capacity mismatch by selectively reducing the capacity devoted to geometry while preserving recognition performance, achieving strong results on PanNuke and Lizard together with zero-shot transfer to MoNuSeg and CPM17. Together with its efficient overall design, LiNA establishes a new Pareto-efficient operating point among the compared joint-analysis systems. More broadly, our results suggest that the accuracy–efficiency trade-off can be improved by selectively reducing capacity in components that are over-provisioned for the task.

AI USE STATEMENT

We used generative AI tools to help implement and debug code, to provide feedback on the research methodology and experiments, and to edit and polish the manuscript. We did not use generative AI tools to generate data or to produce the experimental results reported in this paper. The authors reviewed all AI-assisted suggestions and take full responsibility for the content of the paper.

REPRODUCIBILITY STATEMENT

Full implementation details, including the backbone, the detection and shape heads, prototype distillation, the loss functions, the training schedule, hyperparameter search ranges, and the detection/classification and panoptic-quality scoring protocols, are given in the Appendix. All experiments use the public PanNuke, Lizard, MoNuSeg, and CPM17 datasets, and the split and evaluation protocol for each is specified in the paper.

REFERENCES

Junxiao Chen, Ruixue Wang, Wei Dong, Hua He, and Shiyong Wang. Histonext: dual-mechanism feature pyramid network for cell nuclear segmentation and classification. *BMC Medical Imaging*, 25:9, 2025.

Junzhou Chen, Qian Huang, Yulin Chen, Linyi Qian, and Chengyuan Yu. Enhancing nucleus segmentation with haru-net: a hybrid attention based residual u-blocks network. *arXiv preprint arXiv:2308.03382*, 2023a.

Shengcong Chen, Changxing Ding, Minfeng Liu, Jun Cheng, and Dacheng Tao. Cpp-net: Context-aware polygon proposal network for nucleus segmentation. *IEEE Transactions on Image Processing*, 32:980–994, 2023b.

Gayatri Deshmukh, Onkar Susladkar, Dhruv Makwana, Sparsh Mittal, et al. Feednet: A feature enhanced encoder-decoder lstm network for nuclei instance segmentation for histopathological diagnosis. *Physics in Medicine & Biology*, 67(19):195011, 2022.

Alexandre Filiot, Nicolas Dop, Oussama Tchita, Auriane Riou, Rémy Dubois, Thomas Peeters, Daria Valter, Marin Scalbert, Charlie Saillard, Geneviève Robin, and Antoine Olivier. Distilling foundation models for robust and efficient models in digital pathology. *arXiv preprint arXiv:2501.16239*, 2025.

Jevgenij Gamper, Navid Alemi Koohbanani, Ksenija Benes, Simon Graham, Mostafa Jahanifar, Syed Ali Khurram, Ayesha Azam, Katherine Hewitt, and Nasir Rajpoot. Pannuke dataset extension, insights and baselines. *arXiv preprint arXiv:2003.10778*, 2020.

Simon Graham, Quoc Dang Vu, Shan E Ahmed Raza, Ayesha Azam, Yee Wah Tsang, Jin Tae Kwak, and Nasir Rajpoot. Hover-net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. *Medical image analysis*, 58:101563, 2019.

Liyong Guo et al. Predicting multi-codebook vector quantization indexes for knowledge distillation. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2023.

Dominik Hirling and Peter Horvath. Cell segmentation and representation with shape priors. *Computational and Structural Biotechnology Journal*, 21:742–750, 2023.

Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, et al. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024.

Fabian Hörst, Moritz Rempe, Helmut Becker, Lukas Heine, Julius Keyl, and Jens Kleesiek. Cellvit++: Energy-efficient and adaptive cell segmentation and classification using foundation models. *Computer Methods and Programs in Biomedicine*, 277:109206, 2026.

Jaehoon Jeong et al. IVAAN: Instance-level vision-language alignment via attribute-guided text prompts generation for nuclei analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 29305–29314, 2026.

Solomon Kullback and Richard A Leibler. On information and sufficiency. *The annals of mathematical statistics*, 22(1):79–86, 1951.

Yoseph Linde, Andres Buzo, and Robert Gray. An algorithm for vector quantizer design. *IEEE Transactions on communications*, 28(1):84–95, 1980.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *International Conference on Learning Representations (ICLR)*, 2019.

Saiyang Na, Yuzhi Guo, Feng Jiang, Hehuan Ma, Jean Gao, and Junzhou Huang. Segment any cell: A sam-based auto-prompting fine-tuning framework for nuclei segmentation. *IEEE Transactions on Neural Networks and Learning Systems*, 36(12):19986–19995, 2025.

Zhiliang Peng, Li Dong, Hangbo Bao, Qixiang Ye, and Furu Wei. BEiT v2: Masked image modeling with vector-quantized visual tokenizers. *arXiv preprint arXiv:2208.06366*, 2022.

Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *International conference on medical image computing and computer-assisted intervention*, pp. 265–273. Springer, 2018.

Zhongyi Shui, Yunlong Zhang, Kai Yao, Chenglu Zhu, Sunyi Zheng, Jingxiong Li, Honglin Li, Yuxuan Sun, Ruizhe Guo, and Lin Yang. Unleashing the power of prompt-driven nucleus instance segmentation. In *European conference on computer vision*, pp. 288–304. Springer, 2024a.

Zhongyi Shui, Sunyi Zheng, Chenglu Zhu, Shichuan Zhang, Xiaoxuan Yu, Honglin Li, Jingxiong Li, Pingyi Chen, and Lin Yang. DPA-P2PNet: Deformable proposal-aware P2PNet for accurate point-based cell detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 4864–4872, 2024b.

Oriane Siméoni, Huy V. Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, Francisco Massa, Daniel Haziza, Luca Wehrstedt, Jianyuan Wang, Timothée Darcet, Théo Moutakanni, Leonel Sentana, Claire Roberts, Andrea Vedaldi, Jamie Tolan, John Brandt, Camille Couprie, Julien Mairal, Hervé Jégou, Patrick Labatut, and Piotr Bojanowski. DINOv3, 2025. URL https://arxiv.org/abs/2508.10104.

Qingyu Song, Changan Wang, Zhengkai Jiang, Yabiao Wang, Ying Tai, Chengjie Wang, Jilin Li, Feiyue Huang, and Yang Wu. Rethinking counting and localization in crowds: A purely point-based framework. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3365–3374, 2021.

Cristian Tommasino, Cristiano Russo, and Antonio M Rinaldi. Nulite-lightweight and fast model for nuclei instance segmentation and classification. *Biomedical Signal Processing and Control*, 114:109333, 2026.

Eric Upschulte, Stefan Harmeling, Katrin Amunts, and Timo Dickscheid. Contour proposal networks for biomedical instance segmentation. *Medical Image Analysis*, 77:102371, 2022.

Sanghyun Woo, Shoubhik Debnath, Ronghang Hu, Xinlei Chen, Zhuang Liu, In So Kweon, and Saining Xie. Convnext v2: Co-designing and scaling convnets with masked autoencoders. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 16133–16142, 2023.

Jinlin Yang, Xintao Pang, Chuan Lin, and Tao Tan. Motdnet: Multi organ task decoupling network for cell segmentation. *Medical Image Analysis*, 111:104045, 2026.

Kai Yao, Kaizhu Huang, Jie Sun, and Amir Hussain. Pointnu-net: Keypoint-assisted convolutional neural network for simultaneous multi-tissue histology nuclei segmentation and classification. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 8(1):802–813, 2024.

Ye Zhang, Zijie Fang, Yifeng Wang, Lingbo Zhang, Xianchao Guan, and Yongbing Zhang. Category prompt mamba network for nuclei segmentation and classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 10284–10292, 2025.

Bingchao Zhao, Xin Chen, Zhi Li, Zhiwen Yu, Su Yao, Lixu Yan, Yuqian Wang, Zaiyi Liu, Changhong Liang, and Chu Han. Triple u-net: Hematoxylin-aware nuclei segmentation with progressive dense feature aggregation. *Medical Image Analysis*, 65:101786, 2020.

Bingchao Zhao, Jingxin Luo, Jiatai Lin, Tianpeng Deng, Zaiyi Liu, Guoqiang Han, Ying Wang, and Chu Han. Fkdnuseg: Flawless knowledge distillation for lightweight and fast nuclei instance segmentation and classification. *Medical Image Analysis*, 112:104089, 2026.

Yiqin Zhu, Jianyong Chen, Lingyu Liang, Zhanghui Kuang, Lianwen Jin, and Wayne Zhang. Fourier contour embedding for arbitrary-shaped text detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3123–3131, 2021.

SUPPLEMENTARY MATERIAL

This supplementary material provides the details that were omitted from the main paper for reasons of space, and it is organized into five appendices. Appendix A gives the full implementation and reproducibility details of our method, Appendix B lists the hyperparameter search carried out during development, Appendix C compares against prior Fourier-descriptor methods, Appendix D reports additional morphological characteristics of the nuclei, and Appendix E details the representational-complexity analysis.

A IMPLEMENTATION DETAILS

For completeness, we provide here the implementation details that are needed to reproduce LiNA in full. All of the architectural choices described below are shared across the backbone scales, and the only parts that differ between scales are the backbone itself and the training hyperparameters that are listed at the end of this appendix.

DETECTION

Candidate centers are a fixed regular grid rather than learned queries. Each candidate is refined by two successive offset regressions (center $+ \tanh(\cdot)\, s$ with $s = 6.4$ then $s = 3.2$ pixels). This is a coarse-to-fine cascade, where the first stage moves each anchor close to a nucleus and the second stage makes a small correction. The effective displacement is therefore bounded by the grid spacing, so each candidate stays associated with its own cell. Predictions are matched to ground truth by Hungarian matching. The matching cost is $\mathcal{C} = 0.1\,\|\hat{p} - p\|_2 - \sigma(o)$, where $\hat{p}, p$ are the predicted and ground-truth centers and o is the objectness logit, and pairs farther than 24 pixels apart are forbidden. Each candidate also predicts an objectness score, trained so that matched candidates score high. At inference we keep candidates with objectness above 0.65. We then apply a distance-based non-maximum suppression that greedily suppresses any candidate within 12 pixels of a higher-scoring one, with no top-k cap. Non-maximum suppression is used only at inference, not during matching.

CLASSIFICATION

The classification head predicts $C{+}1$ channels, the C nucleus types of the dataset ($C{=}5$ for PanNuke) plus a background channel, and is trained with a multi-label binary cross-entropy over one-hot targets, rather than softmax cross-entropy, so that the objectness and type predictions share the same sigmoid formulation and an unmatched candidate simply has all type channels off. At inference the type is the arg-max over the C foreground channels.

FOURIER CONTOUR

Ground-truth contours are extracted from each instance mask with external contour tracing (`cv2.findContours`, `RETR_EXTERNAL`) and resampled to 128 points. Because a closed contour has no canonical start point, we align each contour so that the point nearest the 12 o'clock direction from the centroid is index 0, and additionally use a shift-invariant loss that minimizes over all cyclic shifts, which removes the residual start-point ambiguity. The predicted coefficients are decoded in a local frame and translated by the predicted center, so the zeroth coefficient c_0 acts as a learned residual offset that corrects the predicted center rather than an absolute position. Coordinates are kept in raw pixel units with no scale normalization. We use Fourier degree $N{=}15$, giving $2(2N{+}1){=}62$ real outputs decoded into the 128 contour points.

DENSE MASK HEAD

For the capacity ablation (Table 5b, -D), we replace the Fourier contour head and its shape loss with a standard Mask R-CNN-style per-instance mask head and a per-instance mask loss, leaving detection, classification, Hungarian matching, prototype distillation, and the full training recipe unchanged, and train each variant end-to-end from the same initialization. The three backbone levels are projected by $1{\times}1$ convolutions and summed into a stride-4 mask feature ($3{\times}3$ conv, GroupNorm, ReLU, 256 channels). For each detected center a fixed 48 px box is RoIAlign-cropped at

14×14, decoded by four 3×3 conv + GroupNorm + ReLU blocks, upsampled $2\times$ by a transposed convolution to a 28×28 per-instance mask, and read out by a 1×1 convolution. Because one mask is decoded per detected instance, this head's cost scales with the number of nuclei per tile, whereas the Fourier head decodes each contour analytically at negligible cost. Since only the shape representation pathway differs, the dense variant does not improve any evaluated task metric over the Fourier head at any backbone scale (Table 5b).

PROTOTYPE DISTILLATION

The frozen teacher encoder is sampled at each predicted center by bilinear grid sampling. In Phase 1 the codebook $\mathbf{E}$ is learned end-to-end through the classification loss at foreground locations, and the classifier is fed the teacher-derived prototype embedding concatenated with the region-of-interest feature. In Phase 2 (from epoch 70) the teacher is used only to supply the target code distribution under `no_grad`, while the classifier is fed the *student*-derived embedding produced by the code predictor. The student is trained to match the teacher code distribution by the same foreground-masked KL divergence used in the main paper. Inference uses the student path only, so the teacher is never invoked at deployment. The codebook has $K{=}128$ prototypes of dimension $D{=}256$ with softmax temperature $\tau{=}0.07$.

CAB LOSS

The distance-field term uses a precomputed signed distance field of the ground-truth mask, sampled at the predicted contour points and clamped to $[-32, 32]$ pixels. A smooth-L1 penalty toward 0 then drives the absolute value $|\text{SDF}|$ to zero, which matches the point-to-boundary distance $d(c, \mathcal{B})$ in the main text. Predicted centers are detached in this term so it supervises only the Fourier coefficients. The landmark term selects high-curvature landmarks on the ground-truth contour, defined as strict local maxima of the curvature whose value exceeds $1.5\times$ the mean. It keeps the top 8 such points, and pulls each to its nearest predicted contour point with a Huber penalty (weight 0.3).

TRAINING

The student backbone is DINOv3 ConvNeXt-Small on 256×256 inputs, with the frozen H0-mini teacher (86M parameters) Filiot et al. (2025) used only during training. We train for 180 epochs with AdamW (learning rate 1×10^{-4}, weight decay 10^{-4}, batch size 80) on a single NVIDIA A6000 48GB GPU, using a 5-epoch linear warm-up followed by cosine annealing. We use model exponential moving averaging and evaluate every 2 epochs. The CAB loss is enabled from epoch 10, once detection produces reasonable center predictions, and Phase 2 distillation from epoch 70, once the codebook is sufficiently learned. Loss weights are 2.0 (classification), 0.1 (center regression), 0.5 (Fourier), 0.5 (CAB distance field), 0.3 (CAB landmark), and 2.0 (prototype distillation). The seed is set to the fold index, and we report the mean over the three PanNuke folds. To support full reproduction, all configuration files, training and evaluation code, and trained model weights for every reported result will be made publicly available.

BASELINES AND EFFICIENCY MEASUREMENT

For a fair efficiency comparison, all parameter counts, FLOPs, and latencies are measured by us on a single NVIDIA A6000 GPU under identical settings (FP32, batch size 1, and including post-processing). Every method is compared on the same hardware. For accuracy metrics, every experiment we run is scored with the public evaluation code of PromptNucSeg Shui et al. (2024a) under an identical protocol, and for methods whose numbers already exist for the same dataset and protocol we cite the values reported in their papers.

DETECTION AND CLASSIFICATION SCORING

Detection and classification F1 for all experiments we run follow the public evaluation protocol of PromptNucSeg Shui et al. (2024a). For prior methods with published results under the same PanNuke evaluation setting, we report the corresponding published values. For the cross-dataset benchmarks, images are first brought to the same $40\times$ resolution so the matching distance remains

comparable. Panoptic quality is computed per class at IoU > 0.5, and we report avgPQ, the class-averaged PQ obtained as the mean over the five foreground classes, computed identically for every method.

B Hyperparameter Search

We report the hyperparameters explored during development, the range of values tried, the final choice, and the selection criterion (Table S1). Hyperparameters not listed follow Shui et al. (2024a) and are kept fixed, and final values are provided in the released configuration files.

Param	Values tried	Final	Criterion
η	$\{5, 10, 16, 20, 32\} \times 10^{-5}$	1×10^{-4}	val avgPQ
t_{P2}	40, 60, 70	70	Phase 1 val avgPQ
λ_{CAB}	0.5, 1.0	0.5	val avgPQ
λ_{proto}	0.1, 1.0, 2.0	2.0	loss balance
N	5, 10, 15	15	speed + val avgPQ
K	64, 128	128	val avgPQ

Table S1: Hyperparameter search ranges, final values, and selection criteria. η is the learning rate, t_{P2} the Phase-2 start epoch, λ_{CAB} and λ_{proto} the CAB and prototype-distillation loss weights, N the Fourier degree, and K the prototype count.

C Comparison with Prior Fourier-Descriptor Methods

Contour Proposal Networks (CPN) Upschulte et al. (2022) and FourierNet Hirling & Horvath (2023) are the closest structural priors to LiNA. Both represent each instance as a Fourier-descriptor contour, CPN in a single forward pass and FourierNet through a StarDist-derived dense per-location prediction followed by non-maximum suppression. Both, however, perform binary instance segmentation and do not classify nucleus types, so they cannot be placed directly in the multiclass detection, classification, and segmentation tables of the main paper. We therefore report the comparison here. We run CPN both with its original ResNeXt-50 backbone and with the same ConvNeXt-S backbone as LiNA, whereas FourierNet uses a TensorFlow/Keras codebase and cannot be run with our backbone, so we evaluate it in its published configuration. We first compare on the binary task these methods were designed for, and then extend them to the multiclass setting.

Binary Instance Segmentation and Efficiency

Table S2 reports binary instance segmentation on PanNuke test fold 3, scored with the identical metric of Appendix A, reporting binary panoptic quality (bPQ) and the aggregated Jaccard index (AJI) on the class-agnostic instance map. Both prior Fourier methods fall below LiNA's segmentation quality, with bPQ of 0.649 for CPN and 0.602 for FourierNet against 0.673 for LiNA. FourierNet is a much smaller StarDist-lite model, so part of this gap reflects capacity, yet CPN, which matches LiNA in backbone and size, also trails LiNA. CPN is in addition far more expensive, spending 176 GFLOPs because its dense UNet decoder forms contour proposals at every pixel, about 3.7 times the 47.9 GFLOPs of LiNA.

Table S2: Binary instance segmentation on PanNuke test fold 3. Params in M and FLOPs in G. bPQ is binary panoptic quality and AJI is the aggregated Jaccard index, both from the scorer of Appendix A.

Method	Params	FLOPs	bPQ	AJI
CPN (ConvNeXt-S)	69.41	176.0	0.649	0.636
CPN (ResNeXt-50)	160.88	566.9	0.638	0.623
FourierNet	1.42	3.4	0.602	0.597
LiNA (ours)	56.81	47.9	**0.673**	**0.661**

MULTICLASS EXTENSION

Neither method classifies nucleus types, so we extend each with a per-instance type head and train it under the same protocol. Table S3 compares these multiclass variants with LiNA, using the same detection, classification, and avgPQ scorer as the main paper. Extended to the multiclass setting, both methods fall well below LiNA. Their classification F1 of 0.60 for CPN and 0.47 for FourierNet is below LiNA's 0.72, and their avgPQ of 0.363 and 0.307 is below LiNA's 0.451.

Table S3: Multiclass detection, classification, and avgPQ on PanNuke, with a per-instance type head added to each prior method. CPN uses the ConvNeXt-S backbone averaged over three folds, and FourierNet uses its published configuration.

Method	Det-F1	Cls-F1	avgPQ
CPN (multiclass)	0.75	0.60	0.363
FourierNet (+cls head)	0.84	0.47	0.307
LiNA (ours)	**0.88**	**0.72**	**0.451**

These comparisons separate the shared representation from the rest of LiNA. CPN, FourierNet, and LiNA all encode each instance with a low-order Fourier contour, so this compact representation is established prior art rather than our contribution. What distinguishes LiNA is the joint-analysis system built around it, in which the compact geometry is paired with dedicated detection and classification heads and prototype distillation so that reducing shape capacity does not cost recognition accuracy. Consistent with this, both prior methods fall below LiNA on the binary and multiclass tasks, and the remaining components are quantified in the ablation of Table 6, namely the prototype distillation, the curvature-aware boundary loss.

D ADDITIONAL MORPHOLOGICAL CHARACTERISTICS OF NUCLEI

As a complementary observation, we measure per-class shape statistics on the PanNuke ground-truth masks (Table S4). Across all classes the solidity is close to 0.95 and every instance forms a single connected component, so nuclei are compact and close to convex, and even the rare Dead class is not an exception. These properties are consistent with representing each nucleus by a single closed contour rather than a dense per-pixel mask; the low representational complexity of that contour itself is established quantitatively by the reconstruction and boundary-distance analysis of Fig. 4. Figure S1 shows qualitative comparisons against CellViT and PromptNucSeg.

Table S4: Per-class shape statistics of PanNuke ground-truth nuclei. Solidity is near 0.95 and every instance forms a single connected component.

Class	Solidity	# Components
Neoplastic	0.963	1.00
Inflammatory	0.966	1.00
Connective	0.943	1.00
Dead	0.948	1.00
Epithelial	0.964	1.00
All	0.958	1.00

E REPRESENTATIONAL-COMPLEXITY ANALYSIS PROTOCOL AND ADDITIONAL RESULTS

Protocol. For each ground-truth nucleus we extract the external contour of its unsmoothed instance mask, resample it to 128 points, center it at the instance centroid, and fit a degree-30 Fourier series. An order-N reconstruction low-passes these coefficients to order N and rasterizes the resulting closed contour. Region overlap is the IoU of this reconstruction against the raw instance mask, and

boundary distance is the symmetric Chamfer distance to the raw external contour, normalized by the equivalent diameter $\sqrt{4A/\pi}$. We report both on PanNuke, Lizard, MoNuSeg, and CPM17 (Fig. 4). The ellipse-fit prefilter used during training is non-binding on ground truth, excluding no instance on any of the four datasets, and morphological smoothing of the mask does not inflate fidelity, lowering mean IoU by about 0.2 points, so neither step selects for easy shapes.

Spectral concentration. Excluding the translation term, the cumulative energy of the Fourier coefficients reaches 99.7–99.9% by order $N{=}5$ and 99.99% by $N{=}15$ across the four datasets, so almost all coordinate-signal energy lies in the lowest harmonics. This quantifies effective representational complexity rather than an intrinsic dimension.

Hardest shapes. Restricting the analysis to the lowest-solidity quartile, order $N{=}15$ reconstruction still reaches 97.8–98.9% mean IoU with a 10th percentile of 96–98%, so the saturation is not an average-case effect.

Reconstruction ceiling. Even at order 30 the reconstruction reaches 99.2–99.6% IoU rather than 100%. This small residual is dominated by contour extraction, resampling, and rasterization in the analysis pipeline rather than a lack of representational capacity at high Fourier order, consistent with the spectral concentration above.

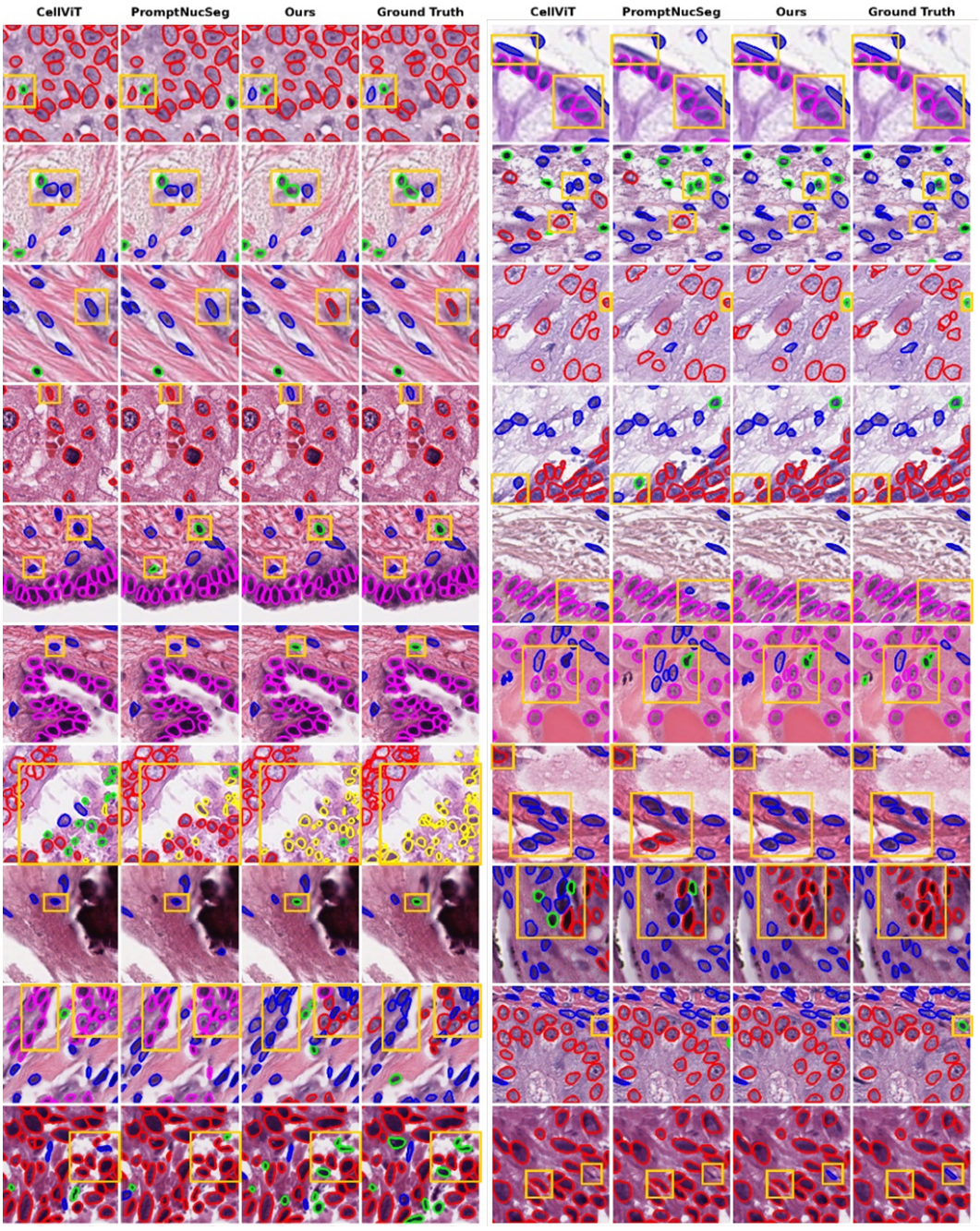


Figure S1: Qualitative comparison on PanNuke. For several tissue tiles we overlay the predicted nucleus contours of CellViT, PromptNucSeg, and LiNA (Ours) against the ground truth, colored by nucleus type. Boxes highlight regions where the methods differ.